BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 13:08:05 ; Search time 50 Seconds
(without alignments)
2831.682 Million cell updates/sec

Title: CAA47749
Perfect score: 4630
Sequence: 1 MASSPAQRRRGNDPLTSSPG.....LFRMNKFSHDLKRKMILQQF 892

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4078.5	88.1	904	23	AAO20516	Protein of APP rel
2	4078.5	88.1	904	24	ABR48218	Human bladder canc
3	4078.5	88.1	904	24	ABU56565	Lung cancer-associ
4	2649	57.2	887	22	ABB63343	Drosophila melanog
5	1872	40.4	864	24	ABJ26013	Aspergillus fumiga
6	1840.5	39.8	785	24	ABJ25413	Aspergillus fumiga
7	1808.5	39.1	903	23	ABP73355	Candida albicans e
8	1073.5	23.2	268	21	AAB56883	Human prostate can
9	902.5	19.5	682	22	AAB62036	P. furiosus mcm po
10	891	19.2	1023	24	ABJ26500	Aspergillus fumiga
11	887	19.2	903	24	ABJ26305	Aspergillus fumiga
12	882.5	19.1	831	24	ABJ25705	Aspergillus fumiga
13	875	18.9	167	24	ABU70820	Human adipocyte Se
14	869	18.8	863	22	AAB95203	Human protein sequ
15	860	18.6	874	24	ABJ25900	Aspergillus fumiga
16	846	18.3	910	23	ABP73643	Candida albicans e
17	835	18.0	901	23	AAG66003	P. patens cell cyc
18	834	18.0	878	23	ABP73678	Candida albicans e
19	834	18.0	878	23	ABG93380	C. albicans BAX-as
20	830.5	17.9	880	23	ABP73688	Candida albicans e
21	819	17.7	819	22	ABB60488	Drosophila melanog
22	816.5	17.6	866	22	ABB58299	Drosophila melanog
23	810	17.5	718	24	ABJ25741	Aspergillus fumiga
24	804	17.4	713	24	ABJ26341	Aspergillus fumiga
25	802	17.3	971	23	ABG93369	S. cerevisiae BAX-
26	800	17.3	733	22	ABB60401	Drosophila melanog
27	783	16.9	724	21	AAB56475	Human prostate can
28	778	16.8	766	21	AAB43937	Human cancer assoc
29	776	16.8	921	24	ABJ25714	Aspergillus fumiga
30	774	16.7	720	23	AAO22751	Maize prolifera ge
31	763	16.5	819	23	AAU69420	Lung small cell ca
32	758.5	16.4	817	22	ABB60368	Drosophila melanog
33	758.5	16.4	840	24	ABU56428	Lung cancer-associ
34	758.5	16.4	841	23	ABB09802	Amino acid sequenc
35	757	16.3	728	23	ABP73710	Candida albicans e
36	754	16.3	957	24	ABJ26314	Aspergillus fumiga
37	753.5	16.3	814	24	ABJ26042	Aspergillus fumiga
38	745.5	16.1	825	23	ABB09803	Amino acid sequenc
39	743	16.0	720	22	ABB61153	Drosophila melanog
40	739.5	16.0	809	23	ABP73372	Candida albicans e
41	732.5	15.8	632	24	ABJ25442	Aspergillus fumiga
42	719	15.5	1049	22	AAB62034	P. furiosus mcm po
43	671	14.5	603	22	AAM94016	Human stomach canc
44	671	14.5	603	22	AAB95303	Human protein sequ
45	650.5	14.0	1115	22	AAB96812	Putative P. abyssi

ALIGNMENTS

RESULT 1
AAO20516
ID AAO20516 standard; Protein; 904 AA.
XX
AC AAO20516;
XX
DT 27-JUN-2002 (first entry)
XX
DE Protein of APP related human homologue hCP50592.
XX
KW Neuroprotective; nootropic; transgenic fly; Alzheimer's disease; Abeta;
KW amyloid precursor protein; tissue-specific expression control; human APP;
KW APP pathway modulator; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200226820-A2.
XX
PD 04-APR-2002.
XX
PF 01-OCT-2001; 2001WO-EP11345.
XX
PR 29-SEP-2000; 2000US-236893P.
PR 14-JUN-2001; 2001US-298309P.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Cohen D, Dengler UJ, Finelli AL, Freuler F, Konsolaki M;
PI Reinhardt MWHM, Zusman S;
XX
DR WPI; 2002-315796/35.

```
DR   N-PSDB; AAK99410.
XX
PT   New transgenic fly, containing DNA encoding an Abeta portion of human
PT   APP, useful for identifying agents which modulate the APP pathway and
PT   which can be used to treat Alzheimer's disease -
XX
PS   Disclosure; Page 123-124; 129pp; English.
XX
CC   The invention relates to a transgenic fly whose genome comprises DNA
CC   encoding a polypeptide having the Abeta portion of human amyloid
CC   precursor protein (APP), fused to a signal sequence. The DNA sequence
CC   encodes a 123 (Abeta40) or 129 (Abeta42) amino acid sequence, given in
CC   the specification. The DNA sequence is operably linked to a tissue-
CC   specific expression control sequence. Expression of the sequence gives
CC   the fly an altered phenotype. The purpose of the invention is for
CC   identifying agents that inhibit or promote the expression and/or function
CC   of genes or encoded polypeptides which modify the APP pathway. The agent
CC   is a compound, triple helix DNA, antisense oligonucleotide, double
CC   stranded RNA molecule, ribozyme, or particularly an antibody. It is used
CC   to treat conditions such as Alzheimer's disease. The agent can be used as
CC   an APP pathway modulator or in gene therapy. This sequence represents the
CC   protein of the APP related human homologue hCP50592.
XX
SQ   Sequence   904 AA;

  Query Match             88.1%;  Score 4078.5;  DB 23;  Length 904;
  Best Local Similarity   89.8%;  Pred. No. 0;
  Matches  817;  Conservative    8;  Mismatches   52;  Indels   33;  Gaps    6;

Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              ||||||||||||||||||||||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTA--------MRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
                 |||    |    :         |: |    |       ||||||||:||||||||||
Db        127 ---GLGRMRRGLLYDSDEEDEERPARKRRQVERA-------TEDGEEDEEMIESIENLED 176

Qy        166 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 236

Qy        226 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 296

Qy        286 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 356

Qy        346 SAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTG 405
              |||||||||||||||||||||||||| |||||||||||||||  ||||||||
Db        357 SAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTG 416

Qy        406 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 476

Qy        466 IFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 525
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        477 IFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 536

Qy        526 EKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 585
              |||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||||
Db        537 EKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 596

Qy        586 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 656

Qy        646 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 716

Qy        706 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGP 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA-- 774

Qy        766 RAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
              ||    | |       :       |    : |:    |||||||||||||||||||||
Db        775 EAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLF 834

Qy        823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 894

Qy        883 LKRKMILQQF 892
              ||||||||||
Db        895 LKRKMILQQF 904


RESULT 2
ABR48218
ID   ABR48218 standard; Protein; 904 AA.
XX
AC   ABR48218;
XX
DT   12-JUN-2003  (first entry)
XX
DE   Human bladder cancer associated protein sequence SEQ ID NO:155.
XX
KW   Human; bladder cancer; cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2003003906-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US21338.
XX
PR   03-JUL-2001; 2001US-302814P.
PR   03-AUG-2001; 2001US-310099P.
PR   08-NOV-2001; 2001US-343705P.
PR   13-NOV-2001; 2001US-350666P.
PR   12-APR-2002; 2002US-372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Mack DH,  Aziz N;
XX
DR   WPI; 2003-201532/19.
DR   N-PSDB; ACC51032.
XX
PT   Detecting a bladder cancer-associated transcript in a cell from a
PT   patient, comprises contacting a biological sample from the patient with
PT   a bladder cancer-associated polynucleotide or antibody  -
XX
PS   Claim 10; Page 283-284; 307pp; English.
XX
CC   The present invention describes a method for detecting a bladder cancer-
CC   associated transcript in a cell from a patient. The method comprises
CC   contacting a biological sample from the patient with a polynucleotide
CC   that selectively hybridises to a sequence that is 80 % identical to a
CC   table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC   encode the human bladder cancer-associated proteins given in ABR48146 to
CC   ABR48242). Bladder cancer-associated sequences from the present invention
CC   have cytostatic activities, and can be used in antisense gene therapy and
CC   in vaccine production. The method can be used for detecting a bladder
CC   cancer-associated transcript in a cell from a patient. The method is
CC   useful in diagnosing or treating bladder cancer and in screening for
CC   compounds that modulate bladder cancer, such as hormones or antibodies.
CC   The nucleic acid molecules from the present invention may be used in
```

```
CC   various screening and diagnostic methods, and for gene therapy, vaccine
CC   and/or antisense/inhibition applications.
XX
SQ   Sequence   904 AA;

  Query Match             88.1%;  Score 4078.5;  DB 24;  Length 904;
  Best Local Similarity   89.8%;  Pred. No. 0;
  Matches  817;  Conservative    8;  Mismatches   52;  Indels   33;  Gaps    6;

Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              ||||||||||||||||||||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTA--------MRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
                 |||    |    :         |: |    |        ||||||||:||||||||||
Db        127 ---GLGRMRRGLLYDSDEEDEERPARKRRQVERA-------TEDGEEDEEMIESIENLED 176

Qy        166 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 236

Qy        226 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 296

Qy        286 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 356

Qy        346 SAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTG 405
              |||||||||||||||||||||||||||||| ||||||||||||||||  ||||||||
Db        357 SAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTG 416

Qy        406 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 476

Qy        466 IFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 525
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        477 IFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 536

Qy        526 EKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 585
              |||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||||
Db        537 EKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 596

Qy        586 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 656

Qy        646 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 716

Qy        706 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGP 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA-- 774

Qy        766 RAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
               ||    | |       :      |    : |:    ||||||||||||||||||||||
Db        775 EAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLF 834

Qy        823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 894

Qy        883 LKRKMILQQF 892
              ||||||||||
Db        895 LKRKMILQQF 904
```

```
RESULT 3
ABU56565
ID   ABU56565 standard; Protein; 904 AA.
XX
AC   ABU56565;
XX
DT   02-APR-2003  (first entry)
XX
DE   Lung cancer-associated polypeptide #158.
XX
KW   Lung cancer-associated polypeptide; cytostatic; emphysema;
KW   antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW   small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW   chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW   interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS   Unidentified.
XX
PN   WO200286443-A2.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US12476.
XX
PR   18-APR-2001; 2001US-284770P.
PR   10-MAY-2001; 2001US-290492P.
PR   09-NOV-2001; 2001US-339245P.
PR   13-NOV-2001; 2001US-350666P.
PR   29-NOV-2001; 2001US-334370P.
PR   12-APR-2002; 2002US-372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Aziz N,  Murray R;
XX
DR   WPI; 2003-093161/08.
DR   N-PSDB; ABX76294.
XX
PT   Detecting a lung cancer-associated transcript in a cell from a patient
PT   for treating lung cancer, by contacting a biological sample from the
PT   patient with a polynucleotide that exhibits increased or decreased
PT   expression in lung cancer  -
XX
PS   Claim 27; Page 310; 453pp; English.
XX
CC   The invention relates to a method for detecting a lung cancer-associated
CC   transcript in a cell from a patient, comprising contacting a biological
CC   sample from the patient with a polynucleotide that selectively hybridises
CC   to a sequence that is at least 80 % identical to a gene that exhibits
CC   increased or decreased expression in lung cancer samples. Lung
CC   cancer-associated polynucleotides and polypeptides are used for
CC   identifying a compound that modulates a lung cancer-associated
CC   polypeptide, for inhibiting proliferation of a lung cancer-associated
CC   cell to treat lung cancer in a patient and for treating a mammal having
CC   lung cancer by administering a modulatory compound identified. The
CC   methods are useful for treating lung cancer, such as small cell lung
CC   cancer, non-small cell lung cancer or other benign or precancerous
CC   lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
CC   pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
CC   pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
CC   and polypeptides are useful for diagnostic purposes and as targets for
CC   screening for therapeutic compounds that modulate lung cancer, such as
CC   antibodies. Sequences ABU56408-ABU56745 represent lung cancer-associated
CC   polypeptides of the invention.
XX
SQ   Sequence   904 AA;

  Query Match             88.1%;  Score 4078.5;  DB 24;  Length 904;
  Best Local Similarity   89.8%;  Pred. No. 0;
  Matches  817;  Conservative    8;  Mismatches   52;  Indels   33;  Gaps    6;

Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
                 ||||||||||||||||||||||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTA--------MRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
                  |||    |    :         |: |    |       ||||||||:||||||||||
Db        127 ---GLGRMRRGLLYDSDEEDEERPARKRRQVERA-------TEDGEEDEEMIESIENLED 176

Qy        166 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 225
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 236

Qy        226 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 285
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 296

Qy        286 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 345
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 356

Qy        346 SAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTG 405
                 |||||||||||||||||||||||||||||| ||||||||||||||||  ||||||||
Db        357 SAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTG 416

Qy        406 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 465
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 476

Qy        466 IFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 525
                 |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        477 IFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 536

Qy        526 EKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 585
                 |||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||||||||
Db        537 EKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 596

Qy        586 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 645
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 656

Qy        646 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 705
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 716

Qy        706 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGP 765
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA-- 774

Qy        766 RAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
                  ||    | |       :       |    : |:    ||||||||||||||||||||||
Db        775 EAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLF 834

Qy        823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 882
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 894

Qy        883 LKRKMILQQF 892
                 ||||||||||
Db        895 LKRKMILQQF 904


RESULT 4
ABB63343
ID   ABB63343 standard; Protein; 887 AA.
XX
AC   ABB63343;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 16821.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07446.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 16821; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   887 AA;

  Query Match             57.2%;  Score 2649;  DB 22;  Length 887;
  Best Local Similarity   57.7%;  Pred. No. 2.8e-229;
  Matches  527;  Conservative  134;  Mismatches  179;  Indels   74;  Gaps   12;

Qy         12 NDPLTSSPGRSSRRTD--ALTSSPGRDLPPFEDESEGLLGTEGPLE--EEEDGEELIGDG 67
              : |  ::|  :: | |  |  :||  |  |||:| | :|| :   :  |||||||| ||
Db          5 SSPPPNTPSDAAERRDLRAAMTSPVGDFEPFENEDE-ILGDQTVRDEAEEEDGEELFGDN 63

Qy         68 MERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCM 127
              || ||| :|||| |:    |  |::|  |::   | ||:
Db         64 MENDYRPMPELDHYD-PALLDDEDDFSEMSQGDRFAAE---------------------- 100

Qy        128 TAMRRTRSALPASAAS---GAGTEDGE--------------------EDEQMIESIENLE 164
              : |||   |           | |  | |                    || :|:|||||||
Db        101 SEMRRRDRAAGIHRDDRDLGFGQSDDEDDVGPRAKRRAGEKAAVGEVEDTEMVESIENLE 160

Qy        165 DLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYE 224
              | |||| :|||| ||| || :||::|||| ||  |   :::||  ||::|  | ||:|
Db        161 DTKGHSTKEWVSMLGPRTEIANRFQSFLRTFVDERGAYTYRDRIRRMCEQNMSSFVVSYT 220

Qy        225 DLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELR 284
              ||| :|||||||||||| ::|:|||: | ::||:::| |:|:|  |||||| |||:|||| |
Db        221 DLANKEHVLAYFLPEAPFQMLEIFDKVAKDMVLSIFPTYERVTTEIHVRISELPLIEELR 280

Qy        285 SLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPEC 344
              : |:||||||:|| |||: |||||||::||:| || :||||| |||| |:||||||||
Db        281 TFRKLHLNQLVRTLGVVTATTGVLPQLSVIKYDCVKCGYVLGPFVQSQNTEIKPGSCPEC 340

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
              || ||| :|||:|:|:|||:| :|||||:: | |:||||| |||||| | |  |||:|:|
Db        341 QSTGPFSINMEQTLYRNYQKITLQESPGRIPAGRIPRSKDVILLADLCDQCKPGDELEVT 400
```

```
Qy        405 GIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGE 464
Db        401 GIYTNNYDGSLNTDQGFPVFATVIIANHVVVKDSKQVVQSLTDEDIATIQKLSKDPRIVE 460

Qy        465 KIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKY 524
Db        461 RVVASMAPSIYGHDYIKRALALALFGGESKNPGEKHKVRGDINLLICGDPGTAKSQFLKY 520

Qy        525 IEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQ 584
Db        521 TEKVAPRAVFTTGQGASAVGLTAYVRRNPVSREWTLEAGALVLADQGVCLIDEFDKMNDQ 580

Qy        585 DRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIIS 644
Db        581 DRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSMTFSENVNLSEPILS 640

Qy        645 RFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPL 704
Db        641 RFDVLCVVKDEFDPMQDQQLAKFVVHSHMKHHPSEEEQPELEE--------PQLKTVDEI 692

Qy        705 PQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGG 764
Db        693 PQDLLRQYIVYAKENIRPKLTNIDEDKIAKMYAQLRQESFATGSLPITVRHIESVIRMS- 751

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAG-----ELHRHTEVQRH---RSMRKTFARYLSFRRDN 816
Db        752 -EAHARMHL-----RENVMEADVSMAIRMMLESFIEAQKFSVMKKMRSTFQKYLSFQKDH 805

Qy        817 NELLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRM 876
Db        806 SELLFFILRQLTLDQLAYIRCKDGPGATHVEIMERDLIERAKQLDIVNLKPFYESDLFRT 865

Qy        877 NKFSHDLKRKMILQ 890
Db        866 NGFSYDPKRRIILQ 879
```

```
RESULT 5
ABJ26013
ID   ABJ26013 standard; Protein; 864 AA.
XX
AC   ABJ26013;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #671.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US13142.
XX
PR   23-APR-2001; 2001US-285697P.
PR   27-APR-2001; 2001US-287066P.
PR   05-JUN-2001; 2001US-295890P.
PR   09-JUL-2001; 2001US-303899P.
PR   31-AUG-2001; 2001US-316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of
PT   Aspergillus fumigatus, useful for treating or preventing infections by
PT   A. fumigatus, or for treating a non-infectious disease in a subject
PT   e.g. cancer  -
XX
PS   Disclosure; Page -; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention.
XX
SQ   Sequence   864 AA;
```

```
  Query Match             40.4%;  Score 1872;  DB 24;  Length 864;
  Best Local Similarity   48.0%;  Pred. No. 3.5e-159;
  Matches  399;  Conservative  110;  Mismatches  213;  Indels  110;  Gaps   12;

Qy         14 PLTSSP----GRSSRR-------TDALTSSPGRDLPPF-------------EDESEGLLG 49
Db          7 PSTSAANRGLGRLNRRKRSREPDDEASSVAPPSSAPPSSPPMLPLDEDDIERDEEAELLG 66

Qy         50 TEGPLE---EEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADG 106
Db         67 DIDDIDEMAEDEDGIDLFGDNFERDYRDVGN-DRYEGEYID-DDGDHEELDIATRRQLEA 124

Qy        107 PCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEE-------------- 152
Db        125 RLN--------------------RRDRELERRRRMPAAFLQDDDEGDMDLTRQPRRRRH 163

Qy        153 ------------DEQMIE-SIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSH 199
Db        164 HYDEDREDIEMGDDAMEELSLEELADVKAANITDWVLQPQVLRSIYREFKAFLTEFTDQA 223

Qy        200 GHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAM 259
Db        224 GQSVYGHKIKTLGEVNSASLEVSYVHLSSTKAALSYFLANEPTEVLKVFDQVALDVTLFH 283

Qy        260 YPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCN 319
Db        284 YPQYQDIHKEIHVRISDLPIVYTLRQLRQQHLNCLVRVSGVVTRRTGVFPQLKYVMFICQ 343

Qy        320 KCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRL 379
Db        344 KCNITLGPFQQEASAEVKISYCQNCQSKGPFTINSEKTVYRNYQKLTLQESPGSVPAGRL 403

Qy        380 PRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK 439
Db        404 PRQREVVLLADLIDSAKPGDEIEVTGIYRNSYDAQLNNKNGFPVFATIIEANHIVKSHDQ 463

Qy        440 VAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGK 499
Db        464 LAGFHLTEEDERQIRALSRDPDIVDKIVRSIAPSIYGHQDVKTAVALSLFGGVSKEAQGK 523
```

```
Qy        500 HKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWT 559
Db        524 MAIRGDINVLLLGDPGTAKSQVLKYVEKTAHRAVFATGQGASAVGLTASVRRDPLTSEWT 583

Qy        560 LEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAA 619
Db        584 LEGGALVLADRGTCLIDEFDKMNDQDRTSIHEAMEQQTISISKAGIVTTLQARCAVVAAA 643

Qy        620 NPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSN 679
Db        644 NPIGGRYNSTIPFSQNVELTEPILSRFDILCVVRDTVDPNEDERLANFVIESHHRANPTR 703

Qy        680 K---------------EEEG---------------LANGSAAEPAMPNTYGVEPLPQE 707
Db        704 PLRDQDGNLVDSEGNRIDEEGYRLDKHGNRLPPTPEEIAKREAAQRKAEEEKEGE-IPQE 762

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESM 759
Db        763 LLRKYILYAREKCHPKLYQIDQDKVARLFADMRRESLATGAYPITVRHLEAI 814
```

```
RESULT 6
ABJ25413
ID   ABJ25413 standard; Protein; 785 AA.
XX
AC   ABJ25413;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #71.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US13142.
XX
PR   23-APR-2001; 2001US-285697P.
PR   27-APR-2001; 2001US-287066P.
PR   05-JUN-2001; 2001US-295890P.
PR   09-JUL-2001; 2001US-303899P.
PR   31-AUG-2001; 2001US-316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of
PT   Aspergillus fumigatus, useful for treating or preventing infections by
PT   A. fumigatus, or for treating a non-infectious disease in a subject
PT   e.g. cancer  -
XX
PS   Disclosure; Page -; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention.
XX
SQ   Sequence   785 AA;
```

```
  Query Match             39.8%;  Score 1840.5;  DB 24;  Length 785;
  Best Local Similarity   49.0%;  Pred. No. 2.1e-156;
  Matches  388;  Conservative  102;  Mismatches  207;  Indels   95;  Gaps   10;

Qy         33 PGRDLPPF---------EDESEGLLGTEGPLE---EEEDGEELIGDGMERDYRAIPELDA 80
Db          5 PPPSSPPMLPLDEDDIERDEEAELLGDIDDIDEMAEDEDGIDLFGDNFERDYRDVGN-DR 63

Qy         81 YEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPAS 140
Db         64 YEGEYID-DDGDHEELDIATRRQLEARLN--------------------RRDRELERRR 101

Qy        141 AASGAGTEDGEE-------------------------DEQMIE-SIENLEDLKGHSVRE 173
Db        102 RMPAAFLQDDDEGDMDLTRQPRRRRHHYDEDREDIEMGDDAMEELSLEELADVKAANITD 161

Qy        174 WVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVL 233
Db        162 WVLQPQVLRSIYREFKAFLTEFTDQAGQSVYGHKIKTLGEVNSASLEVSYVHLSSTKAAL 221

Qy        234 AYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQ 293
Db        222 SYFLANEPTEVLKVFDQVALDVTLFHYPQYQDIHKEIHVRISDLPIVYTLRQLRQQHLNC 281

Qy        294 LIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVN 353
Db        282 LVRVSGVVTRRTGVFPQLKYVMFICQKCNITLGPFQQEASAEVKISYCQNCQSKGPFTIN 341

Qy        354 MEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDG 413
Db        342 SEKTVYRNYQKLTLQESPGSVPAGRLPRQREVVLLADLIDSAKPGDEIEVTGIYRNSYDA 401

Qy        414 SLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPS 473
Db        402 QLNNKNGFPVFATIIEANHIVKSHDQLAGFHLTEEDERQIRALSRDPDIVDKIVRSIAPS 461

Qy        474 IYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAI 533
Db        462 IYGHQDVKTAVALSLFGGVSKEAQGKMAIRGDINVLLLGDPGTAKSQVLKYVEKTAHRAV 521

Qy        534 FTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAM 593
Db        522 FATGQGASAVGLTASVRRDPLTSEWTLEGGALVLADRGTCLIDEFDKMNDQDRTSIHEAM 581

Qy        594 EQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVR 653
Db        582 EQQTISISKAGIVTTLQARCAVVAAANPIGGRYNSTIPFSQNVELTEPILSRFDILCVVR 641

Qy        654 DTVDPVQDEMLARFVVGSHVRHHPSNK---------------EEEG------------- 684
Db        642 DTVDPNEDERLANFVIESHHRANPTRPLRDQDGNLVDSEGNRIDEEGYRLDKHGNRLPPT 701

Qy        685 ---LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK 741
Db        702 PEEIAKREAAQRKAEEEKEGE-IPQELLRKYILYAREKCHPKLYQIDQDKVARLFADMRR 760
```

```
Qy        742 ESMATGSIPITV 753
              ||:|||: ||||
Db        761 ESLATGAYPITV 772


RESULT 7
ABP73355
ID   ABP73355 standard; Protein; 903 AA.
XX
AC   ABP73355;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential protein SEQ ID NO 7192.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US49486.
XX
PR   29-DEC-2000; 2000US-259128P.
PR   20-FEB-2001; 2001US-0792024.
PR   22-AUG-2001; 2001US-314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   N-PSDB; ABZ31905.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele
PT   of a gene and placing other allele of the gene under conditional
PT   expression  -
XX
PS   Claim 44; SEQ ID NO 7192; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans protein used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office.
XX
SQ   Sequence   903 AA;

  Query Match            39.1%;  Score 1808.5;  DB 23;  Length 903;
  Best Local Similarity  45.0%;  Pred. No. 2e-153;
  Matches  393;  Conservative  124;  Mismatches  218;  Indels  139;  Gaps   14;

Qy          1 MASSPA-------QRRRGNDPL----------------TSSP--GRSSRRTDALTSSPG 34
              |:| ||        |||    |                 |||  | |      | |||
Db          1 MSSPPAASSSRDNNRRRKRSPYDDEDDDNNNNDRSSIHQPSSPPLGSSPNPHQHLPSSPA 60

Qy         35 --------------------RDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRA 74
                                  :|| | | :|            ||:|::|: : :|:||||
Db         61 IPFDTALDVDDVEEIVHNDIQDLNPSEPDS-----------EEEGDDLM-ENLEQDYRA 107

Qy         75 IPELDAYE-AEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRT 133
               || | |:  :|   |  | ||: |: |         :   |    |        | |
Db        108 NPEQDHYDLGDGNIDDTGDYEEMDAATRR-------RMINQWNRRDAI----LNNANRSR 156

Qy        134 RSA---------------------LPASAASGAGTEDGEEDEQMIE----------SIE 161
              | |                     ||      .  | ::|: | :           |:|
Db        157 RGAFLDDDDDEDEDDEGEDSMGQFGLPIQRRRRREHHDEDQDDLMDDVEIDPFNEELSLE 216

Qy        162 NLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 221
              :| |:|  |: ||:        |    |:|   : |::| :|:  ::  : : | ||| |
Db        217 SLSDVKAPSITEWILQPAVSRSIARELKSFFLEYTDANGDSVYGNKMRTLGEVNAESLEV 276

Qy        222 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 281
              :|:|||  : :|| ||  :| |:|:|||  |:| |   || | :|   :||||:  | :
Db        277 SYKDLADSKAILALFLATSPEEMLKIFDIVAMEAVELHYPNYSQIHQEVHVRITDFPNIL 336

Qy        282 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC 341
               || ||: :||||:: ||||| ||| || ||::| || ||||: |  | |||   |
Db        337 NLRDLRESNLNQLVKVSGVVTRRTGVFIQLKYVKFDCLKCGVVLGPYVQDSNTEVKISFC 396

Qy        342 PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEI 401
              ||| |||::| |:|:|:||||| :||:|| | | |||| :: |||:|||    |::|
Db        397 TNCQSKGPFKLNSEKTLYRNYQRITLQEAPGTVPAGRLPRHREVILLSDLVDVAKPGEDI 456

Qy        402 ELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVG------ELTDEDVKMITS 455
              |:|||| |||||:|| ||||||||:: || : :|::   :|        |:|:::
Db        457 EVTGIYKNNYDGNLNAKNGFPVFATILEANSIRRKESSAFMGGNNLVNMWTEEEIREFRK 516

Qy        456 LSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPG 515
              || :: | :|| ||:|||||||:|||   | :|||| ||:  ||  :|||||||| ||||
Db        517 LSHEKGIIDKIIASMAPSIYGHKDIKTAIACSLFGGVPKDVNGKLSIRGDINVLLLGDPG 576

Qy        516 TAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLI 575
              ||||| ||| || :|||:| |||||||| :|| |:: |::|||||| |||||||:| |||
Db        577 TAKSQILKYAEKTASRAVFATGQGASAVGLTASVRKDPITREWTLEGGALVLADKGTCLI 636

Qy        576 DEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSEN 635
              |||||||||||||||||||||||||||||:| ||| ||||||| |||| :|   ||
Db        637 DEFDKMNDQDRTSIHEAMEQQSISISKAGIVTTLHARCAVIAAANPNGGRYNSTLPLPEN 696

Qy        636 VDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPA- 694
              |||| ||:|||||:|:||| |:|  || || ||: ||:| ||:|:|:   |||     |
Db        697 VDLTYPILSRFDIMCIVRDLVNPESDERLASFVIDSHMRSHPTNEEDILNANGKGGSDAQ 756

Qy        695 -----MPNTYGVE----------------------PLPQEVLKKYIIYAKERVHPKLN 725
                  | :  ||:                     |:||::| ||| ||: :: |||:
Db        757 DDDENMEDEDGVDQPSAARTRSERIEQLNKQKEQEISPIPQDLLIKYIQYARVKIQPKLH 816

Qy        726 QMDQDKVAKMYSDLRKESMATGSIPITVRHIESM 759
              ||: ||:|::|:|||||:: ||| ||||||:||:
Db        817 QMNMDKLARVYADLRKEAITTGSYPITVRHLESI 850


RESULT 8
AAB56883
ID   AAB56883 standard; Protein; 268 AA.
XX
AC   AAB56883;
XX
DT   13-MAR-2001  (first entry)
```

XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1461.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAF16086.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 1894-1895; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 268 AA;

Query Match 23.2%; Score 1073.5; DB 21; Length 268;
Best Local Similarity 91.7%; Pred. No. 5.8e-88;
Matches 211; Conservative 5; Mismatches 13; Indels 1; Gaps 1;

```
Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 89

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGL-G 119
              ||||||||||||||||||||||||||||||||||||||:||||:     |||   |
Db         90 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRHVTGRLAGAWA 149

Qy        120 ACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAG 179
              |||||||||||||||||||||||| ||||||||:||||||||||||||||||||||||||
Db        150 ACAVGSCMTAMRRTRSALPASAASGAATEDGEEDEEMIESIENLEDLKGHSVREWVSMAG 209

Qy        180 PRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAR 229
              ||||||||||||||||||||||||||||||||||||||||||||||||  ::
Db        210 PRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDTGSQ 259
```

RESULT 9

AAB62036
ID AAB62036 standard; Protein; 682 AA.
XX
AC AAB62036;
XX
DT 14-MAY-2001 (first entry)
XX
DE P. furiosus mcm polypeptide with intein absent.
XX
KW MCM; minichromosome maintenance protein; archaeal polypeptide; PCNA;
KW RFC-P38; RFC-P55; RFA; CDC6; FEN-1; dUTPase; ligase; helicase dna2;
KW nucleic acid amplification; polymerase chain reaction; PCR.
XX
OS Pyrococcus furiosus.
XX
PN WO200109347-A2.
XX
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WO-US20532.
XX
PR 30-JUL-1999; 99US-0146580.
XX
PA (STRA-) STRATAGENE.
XX
PI Hogrefe HH, Cline JM, Hansen CJ, Borns MC;
XX
DR WPI; 2001-182959/18.
DR N-PSDB; AAF57040.
XX
PT Composition for improving nucleic acid polymerase reactions, useful
PT e.g. in synthesis or amplification, contains at least one archaeal
PT accessory protein -
XX
PS Claim 189; Fig 47; 147pp; English.
XX
CC The invention provides a composition (A) for enhancing nucleic acid
CC polymerase reactions that comprises an archaeal MCM (minichromosome
CC maintenance protein) and at least one of the archaeal polypeptides (PCNA,
CC RFC-P38 or -P55, RFA, CDC6, FEN-1, dUTPase, ligase, helicase dna2, or
CC helicases 2-8). (A) And similar compositions containing different
CC combinations of accessory proteins, are used to improve performance of
CC synthesis, amplification, mutagenizing, labeling and detecting reactions,
CC e.g. for gene characterization, cloning, detection of allelic variants,
CC diagnosis and screening for disease, particularly where done by
CC polymerase chain reaction (PCR). Some of the proteins also stabilize
CC duplexes during polymerase reactions or improve exonuclease reactions,
CC for example RFA also improves specificity of nucleic acid/protein
CC interaction and PCNA improves polymerase-mediated repair processes and
CC hybridization reactions. Nucleic acids encoding the archaeal polypeptides
CC are used for recombinant production of proteins, and fragments of the
CC nucleic acid as probes and primers for screening related sequences. The
CC accessory proteins increase accuracy and efficiency of polymerase
CC reactions, allow use of lower denaturation and extension temperatures
CC (possibly isothermal processing), and improve synthesis of long targets.
CC The present sequence represents the amino acid sequence of P. furiosus
CC mcm polypeptide with intein absent.
XX
SQ Sequence 682 AA;

Query Match 19.5%; Score 902.5; DB 22; Length 682;
Best Local Similarity 34.5%; Pred. No. 7.3e-72;
Matches 243; Conservative 137; Mismatches 254; Indels 71; Gaps 21;

```
Qy        181 RLEIHHRFKNFLRTHVDSHGHNVFKERISDMCK-ENRESLVVNYEDLAAREHVLAYFLPE 239
              | |:  || |||| : |  |: |:: :|:|:     : |: :::  | : :  ||: : |
Db          3 REEMIERFANFLREYTDEDGNPVYRGKITDLLTITPKRSVAIDWMHLNSFDSELAHEVIE 62

Qy        240 APAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSG 299
              | | :   ::|   |:   : : |     || |  :||    :: :   |:|:||:  |
Db         63 NPEEGISAAEDAIQIVLREDFQRED--VGKIHARFYNLPETLMVKDIGAEHINKLIQVEG 120

Qy        300 VVTSCTGVLPQLSMVKYNCNKCNF-VLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETI 358
```

```
Db        121 IVTRVGEIKPFVSVAVFVCKDCGHEMIVPQKPYESLE-KVKKCEQCGSKN-IELDVNKSS 178

Qy        359 YQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTA 418

Db        179 FVNFQSFRIQDRPETLKGGEMPRFIDGILLDDIVDVALPGDRVIVTGILRVVLEKREKT- 237

Qy        419 NGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHE 478

Db        238 ---PIFRKILEVNHIEPVSKEIQELEISPEEEQIIKELAKRKDIVDAIVDSIAPAIYGYK 294

Qy        479 DIKRGPALALFGG-EPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTG 537

Db        295 EVKKGIALALFGGVSRKLPDGT-RLRGDIHVLLVGDPGVAKSQILRYVANLAPRAIYTSG 353

Qy        538 QGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQS 597

Db        354 KSSSAAGLTAAAVRDEFTGGWVLEAGALVLADGGYALIDELDKMSDRDRSVIHEALEQQT 413

Qy        598 ISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVD 657

Db        414 ISISKAGITATLNARTTVIAAANPKQGRFNRMKNPFEQIDLPPTLLSRFDLIFVLIDEPD 473

Qy        658 PVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAK 717

Db        474 DKIDSEVARHIL--RVRRGES------------EVVAPK------IPHEILRKYIAYAR 512

Qy        718 ERVHPKLNQMDQDKVAKMYSDLRK-------ESMATGSIPITVRHIESMSHGGGPRAHPS 770

Db        513 KNIHPVISEEAMEEIEKYYVRMRKSVKKTKGEEEGIPPIPITARQLEALIRLS--EAHAR 570

Qy        771 AGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKT------------FARYLSFRRDNNE 818

Db        571 MRLSPIVTREDA--REAIKLMEYT--LKQIAMDETGQIDVTILELGQSARKLS----KIE 622

Qy        819 LLLFILKQLVAEQVTYQRNRFGAQ-QDTIEVPEKDLVDK--ARQI 860

Db        623 KILDIIEKL---QKTSER---GAHVNDILEEAKKAGIEKQEAREI 661


RESULT 10
ABJ26500
ID   ABJ26500 standard; Protein; 1023 AA.
XX
AC   ABJ26500;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #1158.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US13142.
XX
PR   23-APR-2001; 2001US-285697P.
PR   27-APR-2001; 2001US-287066P.
PR   05-JUN-2001; 2001US-295890P.
PR   09-JUL-2001; 2001US-303899P.
PR   31-AUG-2001; 2001US-316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of
PT   Aspergillus fumigatus, useful for treating or preventing infections by
PT   A. fumigatus, or for treating a non-infectious disease in a subject
PT   e.g. cancer  -
XX
PS   Disclosure; Page -; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention.
XX
SQ   Sequence   1023 AA;

  Query Match             19.2%;  Score 891;  DB 24;  Length 1023;
  Best Local Similarity   28.7%;  Pred. No. 1.6e-70;
  Matches  268;  Conservative  140;  Mismatches  304;  Indels  222;  Gaps   29;

Qy          1 MASSPAQRRRG---NDPLTSSPGRSSRRTDALTSSPGRDLPPFEDE-SEGLLGTEGPLEE 56

Db          1 MSSPASSRRRGRPAKDTATSSPARSTRSRQLQTSSP---TPRAADEQSQATPRASRRLR- 56

Qy         57 EEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAAD---------GP 107

Db         57 ---GEAAVPSSSPMFFQSSPSKADSSAE---TPDVRMDEPSSPMRESSTMDEGDRTPRGN 110

Qy        108 CGTVTGSWPGLGACAVGSCMTAMRR--TRSALPASA------------------------ 141

Db        111 APTMRDSSP-IRYMSSSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLH 169

Qy        142 -----------------ASGAGTEDGEEDEQMIESIENL-------EDLKGHSVRE-WVS 176

Db        170 SGGFLSSPNRRRRVFVDANGMPATDG--DPRSDATFSNIHPDTSEAEALGGSSTRVIW-- 225

Qy        177 MAGPRLEIH---HRFKNFL-----RTHVDSHGHNVFKERISDMCKENRE----------- 217

Db        226 --GTNISIQDSMSAFKNFLYNFQTKYRLWAEGATEDETRIMGDSAEEREYISMLSTMRQL 283

Qy        218 ---SLVVNYEDLAAREHVLAYF--LPEAPAELLQIFDEAALEVVLAM------------- 259

Db        284 GVTSLNLDAKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQ 343

Qy        260 --------------YPKYDRITNHIHVRISHLP-LVEE-----------------LRSL 286

Db        344 RNQNHNRGLSSGPAVPSSDALSETGRMPQNEIPDLVGEVETKAFKVLPFGLDSTVNMRDL 403

Qy        287 RQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPE--C 344

Db        404 DPADMDKLVSIKGLVIRTTPIIPDMKEAFFRCQVCNHGV-QVDIDRGKIAEPTECPRPVC 462

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
```

```
Db        463 KERNSMQLIHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVT 522

Qy        405 GIYHNNYDGSLNTANGFPV----------FATVILANHVAKKDNK---------- 439

Db        523 GIFRCN----------PVRVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQEL 571

Qy        440 --VAVGE------LTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGG 491

Db        572 SEQAAGDAEQTRRLTAEEEEKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGG 631

Qy        492 EPK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYV 549

Db        632 TNKTFQKGGNPRYRGDINILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYV 691

Qy        550 QRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSL 609

Db        692 TRDPETRQMVLESGALVLSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTL 751

Qy        610 QARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVV 669

Db        752 NARTSILASANPIGSRYNPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLV 811

Qy        670 GSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQ 729

Db        812 NMYLEDRPEHAAEQ----------EILPIEFLTAYITYAKTKVHPVLTPAAG 853

Qy        730 DKVAKMYSDLRKES----MATGSIPITVRHIESM 759

Db        854 KALSDAYVNMRKLGDDIRSSDRRITATTRQLESM 887
```

```
RESULT 11
ABJ26305
ID   ABJ26305 standard; Protein; 903 AA.
XX
AC   ABJ26305;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #963.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US13142.
XX
PR   23-APR-2001; 2001US-285697P.
PR   27-APR-2001; 2001US-287066P.
PR   05-JUN-2001; 2001US-295890P.
PR   09-JUL-2001; 2001US-303899P.
PR   31-AUG-2001; 2001US-316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of
PT   Aspergillus fumigatus, useful for treating or preventing infections by
PT   A. fumigatus, or for treating a non-infectious disease in a subject
PT   e.g. cancer  -
XX
PS   Disclosure; Page -; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention.
XX
SQ   Sequence   903 AA;
```

```
  Query Match             19.2%;  Score 887;  DB 24;  Length 903;
  Best Local Similarity   33.3%;  Pred. No. 2.9e-70;
  Matches  230;  Conservative  109;  Mismatches  252;  Indels  100;  Gaps   13;

Qy        221 VNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRIT-NHIHVRISHLPL 279

Db         66 VSIDEIRAHNRELADGLLTSPFDYSLAFDRALKEVIKTLPNRPSRETADDVNYYCAYVGA 125

Qy        280 VEEL----RSLRQLHLNQLIRTSGVVTSCTGVLPQ-LSMVKYNCNKCNFVLGPFCQS--- 331

Db        126 FGEFSCNPRTLGSTHLNRMISLEGIVTKCSLVRPKIIQSVHYNERKDRFVARRYRDQTMT 185

Qy        332 ----QNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAIL 387

Db        186 TTGITNMNVYP---QEDDEKNPLITEYGYSTYLDHQTISIQEMPERAPAGQLPRSVDVIL 242

Qy        388 LADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK----VAVG 443

Db        243 DDDLVDSAKPGDRIQLVGIYRSL--GNRNASSGSSTFRTVVMANNIIQLSSKSGGGIAQA 300

Qy        444 ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVR 503

Db        301 TITDTDIRNINKVAKKKNVFELLANSLAPSIYGHDYIKKAILLMLLGGMEKNLDNGTHLR 360

Qy        504 GDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAG 563

Db        361 GDINILMVGDPSTAKSQLLRFVLNTAPLAIATTGRGSSGVGLTAAVTSDKETGERRLEAG 420

Qy        564 ALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIG 623

Db        421 AMVLGDRGVVCIDEFDKMSDVDRVAIHEVMEQQTVTIAKAGIHTSLNARCSVLAAANPIY 480

Qy        624 GRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE-- 681

Db        481 GQYDPHKDPHKNIALPDSLLSRFDLLFVVTDDIEDARDRMVSEHVLRMHRYRQPGTEEGA 540

Qy        682 --------------------------------EEGLANGSAAEPAMPNTYGVEPLPQE 707

Db        541 PVREDLNQTLGVGLEDNQDSNQPTEVYEKFNVMLHAGMANSSR------KGKGIEILSIP 594

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATG---SIPITVRHIESMSHGGG 764

Db        595 FIKKYIQYSKSRIKPVLTKGAADHIVATYSALRNDELSGNQRRTSPITARTLETLIRLS- 653

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFIL 824
```

```
Db        654 -TAHAKARLSNRVE-----------------ERDAKVAESILRFAMFKE---------- 684

Qy        825 KQLVAEQVTYQRNRFGAQQDTIEVPEKDLVD 855
Db        685 ---VLEDERRKRRKVTTFDEDSESSESDSED 712
```

RESULT 12
ABJ25705
ID ABJ25705 standard; Protein; 831 AA.
XX
AC ABJ25705;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #363.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.
XX
SQ Sequence 831 AA;

Query Match 19.1%; Score 882.5; DB 24; Length 831;
Best Local Similarity 33.3%; Pred. No. 6.4e-70;
Matches 230; Conservative 109; Mismatches 251; Indels 101; Gaps 14;

```
Qy        221 VNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRIT-NHIHVRISHLPL 279
Db         28 VSIDEIRAHNRELADLL-TSPFDYSLAFDRALKEVIKTLPNRPSRETADDVNYYCAYVGA 86

Qy        280 VEEL----RSLRQLHLNQLIRTSGVVTSCTGVLPQ-LSMVKYNCNKCNFVLGPFCQS--- 331
Db         87 FGEFSCNPRTLGSTHLNRMISLEGIVTKCSLVRPKIIQSVHYNERKDRFVARRYRDQTMT 146

Qy        332 ----QNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAIL 387
Db        147 TTGITNMNVYP---QEDDEKNPLITEYGYSTYLDHQTISIQEMPERAPAGQLPRSVDVIL 203

Qy        388 LADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK----VAVG 443
Db        204 DDDLVDSAKPGDRIQLVGIYRSL--GNRNASSGSSTFRTVVMANNIIQLSSKSGGGIAQA 261

Qy        444 ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVR 503
Db        262 TITDTDIRNINKVAKKKNVFELLANSLAPSIYGHDYIKKAILLMLLGGMEKNLDNGTHLR 321

Qy        504 GDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAG 563
Db        322 GDINILMVGDPSTAKSQLLRFVLNTAPLAIATTGRGSSGVGLTAAVTSDKETGERRLEAG 381

Qy        564 ALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIG 623
Db        382 AMVLGDRGVVCIDEFDKMSDVDRVAIHEVMEQQTVTIAKAGIHTSLNARCSVLAAANPIY 441

Qy        624 GRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE-- 681
Db        442 GQYDPHKDPHKNIALPDSLLSRFDLLFVVTDDIEDARDRMVSEHVLRMHRYRQPGTEEGA 501

Qy        682 ----------------------------------EEGLANGSAAEPAMPNTYGVEPLPQE 707
Db        502 PVREDLNQTLGVGLEDNQDSNQPTEVYEKFNVMLHAGMANSSR------KGKGIEILSIP 555

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATG---SIPITVRHIESMSHGGG 764
Db        556 FIKKYIQYSKSRIKPVLTKGAADHIVATYSALRNDELSGNQRRTSPITARTLETLIRLS- 614

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFIL 824
Db        615 -TAHAKARLSNRVE-----------------ERDAKVAESILRFAMFKE---------- 645

Qy        825 KQLVAEQVTYQRNRFGAQQDTIEVPEKDLVD 855
Db        646 ---VLEDERRKRRKVTTFDEDSESSESDSED 673
```

RESULT 13
ABU70820
ID ABU70820 standard; Protein; 167 AA.
XX
AC ABU70820;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human adipocyte Selected Interacting domain, SID, #451.
XX
KW Human; prey; adipocyte; SID; selected interacting domain;
KW anorectic; antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS Homo sapiens.
XX

```
PN   WO200286122-A2.
XX
PD   31-OCT-2002.
XX
PF   14-MAR-2002; 2002WO-EP03768.
XX
PR   14-MAR-2001; 2001US-275734P.
XX
PA   (HYBR-) HYBRIGENICS.
XX
PI   Legrain P,  Daviet L;
XX
DR   WPI; 2003-103412/09.
DR   N-PSDB; ACA57364.
XX
PT   New complex between two interacting proteins in adipocyte cells, useful
PT   for identifying selected interacting domains that modulate protein
PT   interactions, or for preventing or treating metabolic disorders such as
PT   obesity or diabetes -
XX
PS   Claim 6; Page 260-261; 382pp; English.
XX
CC   The invention relates to a complex between two interacting proteins in
CC   adipocyte cells, given in the specification. The proteins are identified
CC   by selecting a bait protein from a known adipocyte marker and then
CC   performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC   members of an adipocyte cDNA library. The proteins are designated SID
CC   (RTM) (selected interacting domains) proteins. Also included are a
CC   polynucleotide encoding a polypeptide in the adipocyte cells, a
CC   recombinant host cell expressing at least one of the interacting
CC   polypeptides of the complex, selecting a modulating compound in adipocyte
CC   cells, a SID (RTM)) polypeptide comprising any of the 738 amino acid
CC   sequences given in the specification (including its fragment or variant),
CC   a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC   given in the specification (including its fragment or variant), a vector
CC   comprising the SID (RTM) polynucleotide, a recombinant host cell
CC   comprising the vector, a protein chip comprising the polypeptides and
CC   a record comprising all or part of the data, listed in the specification.
CC   The complex, polypeptides, polynucleotides and compounds are
CC   useful for preventing or treating metabolic disorders such as obesity
CC   or diabetes. The polynucleotides are useful as probes or primers. The
CC   complex is particularly useful for identifying selected interacting
CC   domains (SID (RTM)) for screening drugs that modulate the protein
CC   interaction, thus exhibiting the therapeutic effect. The present
CC   sequence represents a SID (prey) protein of the invention.
XX
SQ   Sequence   167 AA;

  Query Match             18.9%;  Score 875;  DB 24;  Length 167;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-70;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        205 KERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYD 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYD 60

Qy        265 RITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFV 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFV 120

Qy        325 LGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESP 371
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESP 167


RESULT 14
AAB95203
ID   AAB95203 standard; Protein; 863 AA.
XX
AC   AAB95203;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:17301.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17301; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   863 AA;

  Query Match             18.8%;  Score 869;  DB 22;  Length 863;
  Best Local Similarity   30.3%;  Pred. No. 1.1e-68;
  Matches  245;  Conservative  128;  Mismatches  282;  Indels  154;  Gaps   23;

Qy          3 SSPAQRRRGND--------PLTSSPG---RSSRRTDALTSSPGRDLPPFEDESEGLLGTE 51
              |||:|||||  |        |: :|||   :|    | | |||
Db         31 SSPSQRRRGEDSTSTGELQPMPTSPGVDLQSPAAQDVLFSSP------------------ 72

Qy         52 GPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT----ASRREAA--D 105
                                        |::    :  : || :    ||    :|| |
Db         73 ------------------------PQM---HSSAIPLDFDVSSPLTYGTPSSRVEGTPRS 105

Qy        106 GPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
              |  ||     | ||:  |  :              : ||  ||    || :
Db        106 GVRGTPVRQRPDLGSAQKGLQVDLQ----------SDGAAAEDIVASEQSL-------- 146
```

```
Qy        166 LKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVD---SHGHNV--------FKERISDM 211
Db        147 --GQKLVIWGTDVNVAACK----ENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEI 200

Qy        212 CKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNH-I 270
Db        201 NVIGEPFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYP--DSILEHQI 258

Qy        271 HVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 330
Db        259 QVRPFNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTT-RVEM 317

Qy        331 SQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLA- 389
Db        318 DRGRIAEPSVCGRCHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHT--VILFAH 375

Qy        390 -DLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATVILANHVAKKDNKVAV 442
Db        376 NDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVSNVKSVYKTHIDVIHYRKTDAKRLH 430

Qy        443 GELTDED----------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
Db        431 G--LDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGT 488

Qy        493 PK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQ 550
Db        489 RKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVM 548

Qy        551 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 610
Db        549 KDPETRQLVLQTGALVLSDNGICCIDEFDRMNESTRSVLHEVMEQQTLSIAKAGIICQLN 608

Qy        611 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVG 670
Db        609 ARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVA 668

Qy        671 SHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD 730
Db        669 LYYQSEEQAEE-------------------ELLDMAVLKDYIAYAHSTIMPRLSEEASQ 708

Qy        731 KVAKMYSDLRKESMATGSIPITVRHIESM 759
Db        709 ALIEAYVDMRKIGSSRGMVSAYPRQLESL 737
```

RESULT 15
ABJ25900
ID ABJ25900 standard; Protein; 874 AA.
XX
AC ABJ25900;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #558.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.
XX
SQ Sequence 874 AA;

Query Match 18.6%; Score 860; DB 24; Length 874;
Best Local Similarity 29.6%; Pred. No. 7.5e-68;
Matches 251; Conservative 127; Mismatches 269; Indels 200; Gaps 25;

```
Qy          2 ASSPA-----QRRRGNDPLTSS----------PGRS-SRRTD----ALTSSPGRDLPPF 40
Db          3 SSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLHSGGFLSSPNRRRRVF 62

Qy         41 EDESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASR 100
Db         63 VD-ANGMPATDGD---------------------PRSDA-TFSNIHPDTSEAEALGGSS 98

Qy        101 REAADGPCGTVTGSWPGLGACAVGSCMTAMR------RTRSALPASAASGAGTEDGEEDE 154
Db         99 TRVIWGTNISIQDS-----------MSAFKNFLYNFQTKYRLWAEGA----TED--ETR 140

Qy        155 QMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKE 214
Db        141 IMGDSAEE---------REYISM----------------------------LSTMRQL 161

Qy        215 NRESLVVNYEDLAAREHVLAYF--LPEAPAELLQIFDEAALEVVLAMYPK--------YD 264
Db        162 GVTSLNLDAKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQ 221

Qy        265 RITNH----------IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSM 313
Db        222 RNQNHNRGLSSVETKAFKVLPFGLDSTVNMRDLDPADMDKLVSIKGLVIRTTPIIPDMKE 281

Qy        314 VKYNCNKCNFVLGPFCQSQNQEVKPGSCPE--CQSAGPFEVNMEETIYQNYQRIRIQESP 371
Db        282 AFFRCQVCNHGV-QVDIDRGKIAEPTECPRPVCKERNSMQLIHNRCVFADKQVIKLQETP 340
```

```
Qy        372 GKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPV-------- 423
Db        341 DSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCN----------PVRVNPRQRT 389

Qy        424 ----FATVILANHVAKKDNK--------------VAVGE------LTDEDVKMITSLSK 458
Db        390 QKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEEEKIKRTAT 449

Qy        459 DQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGT 516
Db        450 RPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDINILLCGDPST 509

Qy        517 AKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLID 576
Db        510 SKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALVLSDGGICCID 569

Qy        577 EFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENV 636
Db        570 EFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRYNPNLPVPQNI 629

Qy        637 DLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMP 696
Db        630 DLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQ------------- 676

Qy        697 NTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKES----MATGSIPIT 752
Db        677 -----EILPIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITAT 731

Qy        753 VRHIESM 759
Db        732 TRQLESM 738
```

Search completed: September 12, 2003, 13:12:27
Job time : 55 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 13:10:40 ; Search time 26 Seconds
(without alignments)
3299.326 Million cell updates/sec

Title: CAA47749
Perfect score: 4630
Sequence: 1 MASSPAQRRRGNDPLTSSPG.....LFRMNKFSHDLKRKMILQQF 892

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4630	100.0	892	1	S42228	replication licens
2	3941	85.1	904	1	T10067	replication licens
3	3539.5	76.4	886	1	JC5085	replication licens
4	2254	48.7	881	2	T26498	hypothetical prote
5	1951.5	42.1	830	1	B48723	replication licens
6	1945.5	42.0	936	2	E96508	hypothetical prote
7	1812	39.1	868	1	S45757	replication licens
8	913	19.7	686	2	H90226	minichromosome mai
9	909.5	19.6	699	2	B72775	probable DNA repli
10	877	18.9	863	1	S65954	replication licens
11	874.5	18.9	863	1	S64720	replication licens
12	873	18.9	879	2	T41059	replication licens
13	858	18.5	720	2	C84540	probable CDC21 pro
14	856	18.5	862	1	S56766	replication licens
15	855	18.5	776	2	T52118	probable replicati
16	848.5	18.3	933	1	S56050	replication licens
17	843.5	18.2	858	2	T47223	replication licens
18	843	18.2	909	1	S26640	replication licens
19	825.5	17.8	735	1	PC4225	replication licens
20	823.5	17.8	666	1	D69103	DNA helicase (EC 3
21	819.5	17.7	866	1	S59872	replication licens
22	807.5	17.4	796	2	T47222	replication licens
23	802	17.3	971	1	A36376	replication licens
24	796.5	17.2	727	2	G84487	probable DNA repli
25	793.5	17.1	1017	1	S64219	replication licens
26	789	17.0	775	1	A39631	replication licens
27	787.5	17.0	734	1	I38080	replication licens
28	786.5	17.0	892	2	T50339	mis5 protein [impo
29	784	16.9	586	1	E69314	replication licens
30	784	16.9	812	2	T19446	hypothetical prote
31	781	16.9	719	1	S70583	replication licens
32	781	16.9	720	2	T47221	replication licens
33	778	16.8	868	2	T43423	probable replicati
34	777.5	16.8	716	1	T01507	replication licens
35	776.5	16.8	719	1	JC4580	replication licens
36	771	16.7	744	1	I51022	replication licens
37	770.5	16.6	720	1	A48723	replication licens
38	770	16.6	760	2	T39991	minichromosome mai
39	769	16.6	807	1	I51685	replication licens
40	768	16.6	733	1	S56767	replication licens
41	766.5	16.6	808	1	S62594	replication licens
42	761.5	16.4	795	1	S22804	replication licens
43	761	16.4	759	2	T24130	hypothetical prote
44	751	16.2	810	2	H88565	protein ZK632.1 [i
45	748.5	16.2	600	2	S52247	replication licens

ALIGNMENTS

RESULT 1
S42228
replication licensing factor MCM2 - human
N;Alternate names: CDCL1; minichromosome maintenance protein MCM2 homolog; protein BM
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Aug-1999
C;Accession: S42228
R;Todorov, I.T.; Pepperkok, R.; Philipova, R.N.; Kearsey, S.E.; Ansorge, W.; Werner, J. Cell Sci. 107, 253-265, 1994
A;Title: A human nuclear protein with sequence homology to a family of early S phase
A;Reference number: S42228; MUID:94230605; PMID:8175912
A;Accession: S42228
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-892 <TOD>
A;Cross-references: EMBL:X67334; NID:g468703; PIDN:CAA47749.1; PID:g468704
C;Comment: The complex of six MCM proteins is one of several proteins that must be bo phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: GDB:MCM2; CDCL1; D3S3194; cdc19
A;Cross-references: GDB:224876; OMIM:116945
A;Map position: 3q21-3q21
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, a
C;Function:
A;Description: part of the replication licensing system that permits DNA replication
C;Superfamily: replication licensing factor MCM2; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; p
F;283-786/Domain: MCM homology <MCM>
F;318-344/Region: zinc finger CCCC motif

Query Match 100.0%; Score 4630; DB 1; Length 892;
Best Local Similarity 100.0%; Pred. No. 1.3e-285;
Matches 892; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGA 120

Qy        121 CAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGP 180

Qy        181 RLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEA 240

Qy        241 PAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 PAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGV 300

Qy        301 VTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQ 360

Qy        361 NYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANG 420

Qy        421 FPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDI 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 FPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDI 480

Qy        481 KRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 KRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGA 540

Qy        541 SAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISI 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISI 600

Qy        601 SKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQ 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 SKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQ 660

Qy        661 DEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 DEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERV 720

Qy        721 HPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQ 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 HPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQ 780

Qy        781 HGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 HGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFG 840

Qy        841 AQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892


RESULT 2
T10067
replication licensing factor MCM2 - mouse
N;Alternate names: minichromosome maintenance protein MCM2 homolog
C;Species: Mus musculus (house mouse)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C;Accession: T10067
R;Kimura, H.; Ohtomo, T.; Yamaguchi, M.; Ishii, A.; Sugimoto, K.
Genes Cells 1, 977-993, 1996
A;Title: Mouse MCM proteins: complex formation and transportation to the nucleus.
A;Reference number: Z16934; MUID:97224213; PMID:9077461
A;Accession: T10067
A;Molecule type: mRNA
A;Residues: 1-904 <KIM>
A;Cross-references: EMBL:D86725; NID:g2381484; PIDN:BAA22148.1; PID:g2381485
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM2; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;294-797/Domain: MCM homology <MCM>
F;329-355/Region: zinc finger CCCC motif

  Query Match             85.1%;  Score 3941;  DB 1;  Length 904;
  Best Local Similarity   86.1%;  Pred. No. 6.8e-242;
  Matches  787;  Conservative   23;  Mismatches   62;  Indels   42;  Gaps    7;

Qy          2 ASSPA-QRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              |||||  |||| :||||||||||||| |||||||||||||||||||||||||:||||||
Db         10 ASSPARQRRRISDPLTSSPGRSSRRADALTSSPGRDLPPFEDESEGLLGTEGPMEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              |||||||||||| ||||| ||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRPIPELDVYEAEGLALDDEDVEELTASQREAAERTMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTAMRRTRSALPASAASG-----------AGTEDGEEDEQMIESIE 161
                 |||           |  |  |  |:                  ||||||||:||||||
Db        127 ---GLG-----------RMRRGLLYDSSEEDEERPARKRRHVERATEDGEEDEEMIESIE 172

Qy        162 NLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 221
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 NLEDLKGHSVRERVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 232

Qy        222 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 292

Qy        282 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC 341
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        293 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCSKCNFVLGPFCQSQNQEVKPGSC 352

Qy        342 PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEI 401
              ||||||||||:|||||||||||||||||||||||| ||||||||||||||||  ||||
Db        353 PECQSAGPFEINMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEI 412

Qy        402 ELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQ 461
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        413 ELTGIYHNNYDGSLNTANGFPVFATIILANHVAKKDNKVAVGELTDEDVKMITGLSKDQQ 472

Qy        462 IGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQF 521
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        473 IGEKIFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQF 532

Qy        522 LKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKM 581
              |||||||||||||||||||| :|||||||||||||||||||||||||||||||||||||||
Db        533 LKYIEKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKM 592

Qy        582 NDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEP 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 NDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEP 652

Qy        642 IISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGV 701
              ||||||:|||||||||||||||||||||||||||||::||| ||   ||||||||||
Db        653 IISRFDVLCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKKDEGLTNGGTLEPAMPNTYGV 712

Qy        702 EPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSH 761
              |||||||||||||||||| ||||||||||:|||||||||||||||||||||||||||||||
Db        713 EPLPQEVLKKYIIYAKERVRPKLNQMDQDKVARMYSDLRKESMATGSIPITVRHIESMIR 772

Qy        762 GGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNE 818
                  ||    | |      :     :    : |:    |||||||||||||||||:
Db        773 MA--EAHARMHLRDYVMEDDVNMAIRVMMESFIDTQKFSVMRSMRKTFARYLSFRRDNND 830

Qy        819 LLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNK 878
              |||||||||||||||||||||||||:|||||:|||||||||||||||:|| ||
Db        831 LLLFILKQLVAEQVTYQRNRFGAQQDTIEIPEKDLMDKARQINIHNLSAFYDSDLFKFNK 890

Qy        879 FSHDLKRKMILQQF 892
              || |||||:|||||
Db        891 FSRDLKRKLILQQF 904


RESULT 3
JC5085
replication licensing factor MCM2 [validated] - African clawed frog
N;Alternate names: MCM2p; minichromosome maintenance protein MCM2 homolog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 01-Dec-2000
C;Accession: JC5085; T47225
R;Miyake, S.; Saito, I.; Kobayashi, H.; Yamashita, S.
Gene 175, 71-75, 1996
```

```
A;Title: Identification of two Xenopus laevis genes, xMCM2 and xCDC46, with sequence hom
A;Reference number: JC5085; MUID:97074651; PMID:8917078
A;Contents: oocyte
A;Accession: JC5085
A;Molecule type: mRNA
A;Residues: 1-886 <MIY>
A;Cross-references: DDBJ:D63919; NID:g1753192; PIDN:BAA09948.1; PID:g1753193
R;Kubota, Y.; Mimura, S.; Nishimoto, S.; Masuda, T.; Nojima, H.; Takisawa, H.
EMBO J. 16, 3320-3331, 1997
A;Title: Licensing of DNA replication by a multi-protein complex of MCM/P1 proteins in X
A;Reference number: Z24400; MUID:97357318; PMID:9214647
A;Accession: T47225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-168,'K',170-438,'G',440-562,'L',564-727,'M',729-886 <KUB>
A;Cross-references: EMBL:U44047; NID:g2231168; PIDN:AAC60223.1; PID:g2231169
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: MCM2
C;Complex: The predominant form is a heterohexamer; the complex consists of MCM2 (PIR:T4
ated, MUID:97357318], sometimes called the RLF-M component of replication licensing fact
C;Function:
A;Description: MCM2 is a component of the replication licensing factor that permits DNA
C;Superfamily: replication licensing factor MCM2; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;279-779/Domain: MCM homology <MCM>
F;314-340/Region: zinc finger CCCC motif

  Query Match             76.4%;  Score 3539.5;  DB 1;  Length 886;
  Best Local Similarity   78.3%;  Pred. No. 1.9e-216;
  Matches  704;  Conservative   57;  Mismatches   95;  Indels   43;  Gaps    8;

Qy         15 LTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRA 74
              : :||   ||| |||||||||||||||||||||: |     |||||||||||| ||||||
Db         10 IATSPRAGSRR-DALTSSPGRDLPPFEDESEGMFGDGVVPEEEEDGEELIGDAMERDYRP 68

Qy         75 IPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTGSWPGLGACAVGSCM 127
              | ||| || |||  |:||||:||||:||||       |   |   |
Db         69 ISELDRYEVEGLD-DEEDVEDLTASQREAAEQSMRMRDREMGRELG-------------- 113

Qy        128 TAMRR-----------TRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVS 176
                |||            | |     |  |     ||||:|||||||||:|||:||||||
Db        114 -RMRRGLLYDSDEEEEDRPARKRRMAERAAEGAPEEDEEMIESIENLEDMKGHTVREWVS 172

Qy        177 MAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYF 236
              ||  ||||:|||||||||||| |||||||:|||||||:||| ||||||||||||||||
Db        173 MAATRLEIYHRFKNFLRTHVDEHGHNVFKEKISDMCKENKESLPVNYEDLAAREHVLAYF 232

Qy        237 LPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIR 296
              |||||||:|:|||||| |||| ||||||||   |||||||||||||||||||||||||
Db        233 LPEAPAEMLKIFDEAAKEVVLVMYPKYDRIAREIHVRISHLPLVEELRSLRQLHLNQLIR 292

Qy        297 TSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEE 356
              |||||| |||||||||||||||||:|||| |||||||:|||||||| |||||:||||
Db        293 TSGVVTCCTGVLPQLSMVKYNCNKCNFILGPFFQSQNQEVRPGSCPECQSFGPFEINMEE 352

Qy        357 TIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLN 416
              |:||||||| |||||||||| ||||||||||||||||  |||||||||||||||||||||
Db        353 TVYQNYQRITIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTGIYHNNYDGSLN 412

Qy        417 TANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYG 476
              |||||||||||||||: |||:|||| |||||||| | :||||::|||:|||||||||||
Db        413 TANGFPVFATVILANHITKKDDKVAVRELTDEDVKAIVALSKDERIGERIFASIAPSIYG 472

Qy        477 HEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTT 536
              ||||||| |||||||| ||||||||||||||||||||||||||||:|||:|||:|||||
Db        473 HEDIKRGLALALFGGEAKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYVEKVASRAVFTT 532

Qy        537 GQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQ 596
              ||||||| :|||||||::|||||||| |||||||||||||||||||||||||||||||
Db        533 GQGASAVGLTAYVQRHPVTKEWTLEAGALVFADRGVCLIDEFDKMNDQDRTSIHEAMEQQ 592

Qy        597 SISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTV 656
              ||||||||||||||||||||:|||||||||||||||||||||:||||||||||||
Db        593 SISISKAGIVTSLQARCTVIAASNPIGGRYDPSLTFSENVDLTEPIVSRFDILCVVRDTV 652

Qy        657 DPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYA 716
              |||||||||||| ||::||||:|:   :||| ||| |:|||:||| ||||||||:||
Db        653 DPVQDEMLARFVVSSHIKHHPSSKD---IANGDAAEFALPNTFGVEALPQEVLKKYIMYA 709

Qy        717 KERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDR 776
              ||:: |||||||||||:|||||||||||||||||||||||        ||    | |
Db        710 KEKIRPKLNQMDQDKVAKIYSDLRKESMATGSIPITVRHIESMIRMA--EAHARMHLRDY 767

Qy        777 RRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLVAEQVT 833
                    :      |    : |:    ||||||||||:||||||||||:||||:|||||
Db        768 VVEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLAFRRDNNELLLFVLKQLIAEQVT 827

Qy        834 YQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892
              |||||:|||||||||||||||||||||||||||||||:|||:||||:||:|:|:|:|||
Db        828 YQRNRYGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSDLFKMNKFTHDVKKKLIIQQF 886


RESULT 4
T26498
hypothetical protein Y17G7B.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
C;Accession: T26498
R;Smye, R.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z20225
A;Accession: T26498
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-881 <WIL>
A;Cross-references: EMBL:AL023828; PIDN:CAA19452.1; GSPDB:GN00020; CESP:Y17G7B.5
A;Experimental source: clone Y17G7B
C;Genetics:
A;Gene: CESP:Y17G7B.5
A;Map position: 2
A;Introns: 55/2; 343/3; 758/2
C;Superfamily: replication licensing factor MCM2; MCM homology
F;270-771/Domain: MCM homology <MCM>

  Query Match             48.7%;  Score 2254;  DB 2;  Length 881;
  Best Local Similarity   52.4%;  Pred. No. 6.8e-135;
  Matches  465;  Conservative  129;  Mismatches  226;  Indels   68;  Gaps   15;

Qy         41 EDESEG---LLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT 97
              :|: :|   :   |    |:|:|| | || ||||||  |||| |   |:  |  ||  |:
Db         23 DDDVDGIDEMFNNEDEDPEDEEGENLFGDDMERDYREQPELDQYSESGMD-DASDVGSLS 81

Qy         98 ASRREAA--------------------DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALP 138
               | | ||                    ||    |     | |           || |
Db         82 VSARRAAEREMAQRDQLLDDDALMYEDGDSEEVDTRRRGRG----------RRGR---- 126

Qy        139 ASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRT-HVD 197
                    |  :|    |:    :: ||:::| ::|: ||      ||  |||||||: |
Db        127 ----GDAADDDSVPMEEEDIPVDILENIRGRTIRDHVSDEAVAKEIERRFKNFLRSFHEP 182

Qy        198 SHGHNVFKERISDMCKENRESLVVNYEDLA--AREHVLAYFLPEAPAELLQIFDEAALEV 255
              :    : : |  |  :|||| |:: ||:    |  ::|||||| |:| | | || ||
Db        183 GNKQTKYIQMIKSMAADNRESLEVSFTDLSDDNGEQNISYFLPEAPNEMLAIMDRAATEV 242

Qy        256 VLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVK 315
              |: ||| | |: | | |||| ||: |::| |||:||| ||||:|||  :|:|||::||
Db        243 VMNMYPFYSRVCNEIKVRISQLPVEEDIRMLRQVHLNMLIRTAGVVTIASGILPQLAVVK 302

Qy        316 YNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVA 375
              |:|  | ::|||| |  ::||:|  || ||  ||||:|:| |:| ||||| :|||| |||
Db        303 YDCVACGYLLGPFVQQNDEEVRPTICPSCQGKGPFELNVENTVYHNYQRITMQESPNKVA 362

Qy        376 ARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAK 435
              | ||||||| ||| || |||  |||||:||:| ||:|||||   ||||| |:| |||:
```

```
Db        363 AGRLPRSKDVILLGDLCDSCKPGDEIEVTGVYTNNFDGSLNYKQGFPVFNTLIHANHITN 422

Qy        436 KDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKN 495
              || |:|  :|||||:| |  ||:|  | :::|:||||||||||:|:||  ||||| || ||
Db        423 KD-KMASDQLTDEDIKAIRELSQDPNISQRVFSSIAPSIYGHDDVKRAIALALFRGEAKN 481

Qy        496 PGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVS 555
              || ||::|||||||||||||||||:|   :: |:: ||||||||| :|||||||||:
Db        482 PGAKHRLRGDINVLLCGDPGTAKSQFLRYAAHIAPRSVLTTGQGASAVGLTAYVQRHPVT 541

Qy        556 REWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTV 615
              |||||||||:||||:||||||||||:|||||||||||||||||||||||||| |||||
Db        542 REWTLEAGAMVLADKGVCLIDEFDKMSDQDRTSIHEAMEQQSISISKAGIVTSLHARCTV 601

Qy        616 IAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRH 675
              |||:|||||||:|: ||:|||||||||:||||:|||:||:|| |:|| ||:||||:|  |
Db        602 IAASNPIGGRYNPTRTFAENVDLTEPILSRFDVLCVIRDSVDSVEDERLAKFVVGNHRTH 661

Qy        676 HPSNKE--EEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVA 733
              ||  |:  :||       |  |    ||  :||::|:|||||:|: || | :   :| :
Db        662 HPDAKKIVKEG---DELEEDQMDERTGVRLIPQDLLRKYIIYARKECHPTLPEQHSEKFS 718

Qy        734 KMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE---- 789
              ::: :|||||||||: ||||:|||       ||    |     | : :  |
Db        719 NIFAQMRKESMATGSVAITVRHVESMIRLS--EAHAKLHL-----RSYVNDEDCAAAIRV 771

Qy        790 -LHRHTEVQR---HRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTY--QRNRFGAQQ 843
              |      |:    | |:|||:|:|:  |  ||||||||||: :|: |   |   |
Db        772 MLESFVNTQKASIMRMMKKTFSRHLTENRSANELLLFILKQLIRQQMHYATARAAAGTIL 831

Qy        844 DTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQ 891
              :: :|| : ::||:|: | |:  || ||:|  | | :|  :| |:|:
Db        832 QSVTIPESEFIEKAQQLRIENVKPFYTSEIFASNNFLYDPSKKTIVQE 879
```

RESULT 5
B48723
replication licensing factor MCM2 homolog nda1 - fission yeast (Schizosaccharomyces pomb
N;Alternate names: cell division cycle protein cdc19
C;Species: Schizosaccharomyces pombe
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 18-Jul-2001
C;Accession: B48723; T40418; T52476
R;Miyake, S.; Okishio, N.; Samejima, I.; Hiraoka, Y.; Toda, T.; Saitoh, I.; Yanagida, M.
Mol. Biol. Cell 4, 1003-1015, 1993
A;Title: Fission yeast genes nda1(+) and nda4(+), mutations of which lead to S-phase blo
A;Reference number: A48723; MUID:94129084; PMID:8298187
A;Accession: B48723
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-830 <MIY>
A;Cross-references: GB:S68468; NID:g545212; PIDN:AAC60569.1; PID:g545213
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Brown, S.; Harris, D.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21927
A;Accession: T40418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-830 <MCD>
A;Cross-references: EMBL:AL121863; PIDN:CAB58403.1; GSPDB:GN00067; SPDB:SPBC4.04c
A;Experimental source: strain 972h-; cosmid c4
R;Forsburg, S.L.; Nurse, P.
J. Cell Sci. 107, 2779-2788, 1994
A;Title: The fission yeast cdc19+ gene encodes a member of the MCM family of replication
A;Reference number: Z26087; MUID:95181580; PMID:7876346
A;Accession: T52476
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-830 <FOR>
A;Cross-references: EMBL:U08048; PIDN:AAC48930.1
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: nda1+
A;Map position: 2
A;Introns: 78/2
A;Note: cdc19+
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, a
C;Function:
A;Description: part of the replication licensing system that permits DNA replication
lidated, MUID:95181580]
C;Superfamily: replication licensing factor MCM2; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; p
F;299-804/Domain: MCM homology <MCM>
F;334-360/Region: zinc finger CCCC motif

  Query Match   42.1%; Score 1951.5; DB 1; Length 830;
  Best Local Similarity 48.2%; Pred. No. 9.7e-116;
  Matches 418; Conservative 116; Mismatches 238; Indels   95; Gaps   12;

```
Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLE----- 55
              | |   : || :: |      ||     |:  |    | | ||:   |  |  :|
Db          1 MDSFRKRGRRDSESLPFESENSSLGATPLSLPPSSPPPEFSDEAAEAL-VEEDIEDLDGE 59

Qy         56 ----EEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTV 111
                 |:|:||:| |:|||||:   ||| |: | |  || |:|||    | | |
Db         60 ALDVEDEEGEDLFGEGMERDYQQNLELDRYDIEELD-DDNDLEELDIGARRAVD------ 112

Qy        112 TGSWPGLGACAVGSCMTAMRRTRSALPASAA----------------SGAGT-------- 147
                                  :||    | |:|                |  ||
Db        113 ----------------ARLRRRDIELDAAAGRTKPAAFLQDEDDDLDSNLGTGFTRHRHR 156

Qy        148 -------EDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHG 200
                    |  ||     :|:: |:|  |: |||::   |  |   |||||  : | :|
Db        157 IYDEYSPNVGALDESGELPLESIADVKADSIAEWVTLDPVRRTIAREFKNFLLEYTDENG 216

Qy        201 HNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMY 260
               :|:  ||  : : | |||:|||  |   : :|||||  ||| : :|||  |||  |  |
Db        217 TSVYGNRIRTLGEVNAESLMVNYAHLGESKPILAYFLANAPAPIFRIFDRVALEATLLHY 276

Qy        261 PKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNK 320
              | |:|| : |||||::||    || ||| ||| |:| |||||  ||: |||  ::: | |
Db        277 PDYERIHSDIHVRITNLPTCFTLRDLRQSHLNCLVRVSGVVTRRTGLFPQLKYIRFTCTK 336

Qy        321 CNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLP 380
              |   |||| |  : |||   |  | | ||| :| | |:| ||||| :|||||  | : |||
Db        337 CGATLGPFFQDSSVEVKISFCHNCSSRGPFVINSERTVYNNYQRITLQESPGTVPSGRLP 396

Qy        381 RSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAK----- 435
              | :: |||||||    |:||::|||| ||:| |||| ||||||||:| |||:::
Db        397 RHREVILLADLVDVAKPGEEIDVTGIYRNNFDASLNTKNGFPVFATIIEANHISQLDGSG 456

Qy        436 -KDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPK 494
               |:  ::  |||:: : | :|:|   |  :| ||:|||||||  ||   | |||||
Db        457 NTDDDFSLSRLTDDEEREIRALAKSPDIHNRIIASMAPSIYGHRSIKTAIAAALFGGVPK 516

Qy        495 NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPV 554
              |  ||||:|||||||| ||||||||||||||:|| : ||:| ||||||| :|| |:: |:
Db        517 NINGKHKIRGDINVLLLGDPGTAKSQFLKYVEKTAHRAVFATGQGASAVGLTASVRKDPI 576

Qy        555 SREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCT 614
              : ||||| |||||||:|||||||||||||||||||||||||||||||||||||:||||||
Db        577 TNEWTLEGGALVLADKGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTTLQARCT 636

Qy        615 VIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVR 674
              :||||||||||: :: |::||:|||||:|||||| ||:|||:|  || || ||| ||:|
Db        637 IIAAANPIGGRYNTTIPFNQNVELTEPILSRFDILQVVKDTVNPEIDEQLANFVVSSHIR 696

Qy        675 HHPSNKEEEGLANGSAAEPAM------PNTYGVE--PLPQEVLKKYIIYAKERVHPKLNQ 726
               ||            | :| |      |   |::  |:||:?|:||| :|:|:| |:| |
Db        697 SHP------------AFDPNMDVLKKVPTETGIDAKPIPQDLLRKYIHFAREKVFPRLQQ 744

Qy        727 MDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRD 786
              ||::|::::|||:|:||:|||| |||||:||      |     | :  |  |
Db        745 MDEEKISRLYSDMRRESLATGSYPITVRHLESAIRLS--EAFAKMQLSEFVRPSHIDKAI 802

Qy        787 AGELHRHTEVQR---HRSMRKTFARYL 810
```

```
                :      |:    ||: :|||:||
Db        803 QVIIDSFVNAQKMSVKRSLSRTFAKYL 829


RESULT 6
E96508
hypothetical protein T12C22.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: E96508
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96508
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-936 <STO>
A;Cross-references: GB:AE005173; NID:g8656002; PIDN:AAF78275.1; GSPDB:GN00141
C;Genetics:
A;Gene: T12C22.19
A;Map position: 1
C;Superfamily: replication licensing factor MCM2; MCM homology

  Query Match             42.0%;  Score 1945.5;  DB 2;  Length 936;
  Best Local Similarity   45.3%;  Pred. No. 2.8e-115;
  Matches  430;  Conservative  150;  Mismatches  268;  Indels  101;  Gaps   20;

Qy         12 NDPLT-SSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGT--------EGPLEEEEDGEE 62
              |:| : :||  :   || |  |  ::   | || |  : |        |   ||||:||:
Db          8 NEPSSPASPSSAGFNTDQLPISTSQNSENFSDEEEAAVDTQVIRDEPDEAEDEEEEEGED 67

Qy         63 LIGDGMERDYRAIPELDAYEAEGL--ALDDE-DVEELTASRREAADGPC----------- 108
              |  |    ||| : | | ||: |:  ::||| |: :    || |||
Db         68 LFNDTFMNDYRKMDENDQYESNGIDDSVDDERDLGQAMLDRR-AADADLDARENRLANRK 126

Qy        109 ------GTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDG---------EED 153
                   :  |               :|:|: :|       |  ||         : |
Db        127 LPHLLHDNDSDDWN----------YRPSKRSRTTVP--PRGNGGDPDGNPPSSPGVSQPD 174

Qy        154 EQMIESIENLED--------------LKGHSVREWVSMAGPRLEIHHRFKNFLRTHV--- 196
              | :  :: :|              ::| ::||||     |  |  :||:|| |:|
Db        175 ISMTDQTDDYQDEDDNDDEAEFEMYRIQG-TLREWVMRDEVRRFIAKKFKDFLLTYVKPK 233

Qy        197 DSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVV 256
              : :|   :   |::|   |: || ::|::       :| :| :||  :|:: :| : :|:
Db        234 NENGDIEYVRLINEMVSANKCSLEIDYKEFIHVHPNIAIWLADAPQPVLEVMEEVSEKVI 293

Qy        257 LAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKY 316
              ::| |  |   |:||:::||: :::|::||:||| :||  ||||  :|| |||  |||
Db        294 FDLHPNYKNIHTKIYVRVTNLPVNDQIRNIRQIHLNTMIRIGGVVTRRSGVFPQLQQVKY 353

Qy        317 NCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAA 376
              :|||  ||||| |:   ||| ||| |||| ||| ||:|:|||:|||:: ||||||  | |
Db        354 DCNKCGAVLGPFFQNSYSEVKVGSCSECQSKGPFTVNVEQTIYRNYQKLTIQESPGTVPA 413

Qy        377 RRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKK 436
              |||| |: ||| ||:|    |:|||:|||| ||:| |||| |||||||||: ||:| ||
Db        414 GRLPRHKEVILLNDLIDCARPGEEIEVTGIYTNNFDLSLNTKNGFPVFATVVEANYVTKK 473

Qy        437 DNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNP 496
              :  :  :|| ||   |  |||| :| |:|  |||||||||||||   |||:|||: ||
Db        474 QDLFSAYKLTQEDKTQIEELSKDPRIVERIIKSIAPSIYGHEDIKTALALAMFGGQEKNI 533

Qy        497 GGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSR 556
              |||::|||||||| ||||||||||||||:||   ||::|||:|||||| :|| | : ||:|
Db        534 KGKHRLRGDINVLLLGDPGTAKSQFLKYVEKTGQRAVYTTGKGASAVGLTAAVHKDPVTR 593

Qy        557 EWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVI 616
              ||||| |||||||||:|||||||||||||| ||||||||||||||||||||||||:||
Db        594 EWTLEGGALVLADRGICLIDEFDKMNDQDRVSIHEAMEQQSISISKAGIVTSLQARCSVI 653

Qy        617 AAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHH 676
              |||||:||||| | :|::||:||:||:|||||||||:| |||| ||||| ||| || :
Db        654 AAANPVGGRYDSSKSFAQNVELTDPILSRFDILCVVKDVVDPVTDEMLAEFVVNSHFKSQ 713

Qy        677 P------SNKEEEGLANGS-AAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQ 729
              |       :  |:|:   | : :|        | ||| :|||: |:|  | ||| ::|
Db        714 PKGGKMEDSDPEDGIQGSSGSTDP--------EVLPQNLLKKYLTYSKLYVFPKLGELDA 765

Qy        730 DKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE 789
              `|:  :|::||:|||    : |  ||:|||       ||    |     ||:    |
Db        766 KKLETVYANLRRESMNGQGVSIATRHLESMIRMS--EAHARMHL-----RQYVTEEDVNM 818

Qy        790 -----LHRHTEVQR---HRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFGA 841
                    |       |:    |::|::| ||:::::| | ||| :||:||   : ::    |:
Db        819 AIRVLLDSFISTQKFGVQRTLRESFKRYITYKKDFNSLLLVLLKELVKNALKFEEIISGS 878

Qy        842 QQ--DTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMI 888
                  ||||  ::|  ||:: :| :|  |: |  |    |  |  | ||
Db        879 NSGLPTIEVKIEELQTKAKEYDIADLRPFFSSTDFSKAHFELDHGRGMI 927


RESULT 7
S45757
replication licensing factor MCM2 [validated] - yeast (Saccharomyces cerevisiae)
N;Alternate names: minichromosome maintenance protein MCM2; protein YBL023c; protein
C;Species: Saccharomyces cerevisiae
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C;Accession: S45757; S50306; S43939; B39630; A39630
R;Goffeau, A.; Jonniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck,
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45745
A;Accession: S45757
A;Molecule type: DNA
A;Residues: 1-868 <GOF>
A;Cross-references: EMBL:Z35784; NID:g536020; PIDN:CAA84842.1; PID:g536021; GSPDB:GN0
R;van Dyck, L.; Jonniaux, J.L.; de Melo Barreiros, T.; Kleine, K.; Goffeau, A.
Yeast 10, 1663-1673, 1994
A;Title: Analysis of a 17.4 kb DNA segment of yeast chromosome II encompassing the ri
human proliferation-associated p120 antigen.
A;Reference number: S50299; MUID:95242843; PMID:7725803
A;Accession: S50306
A;Molecule type: DNA
A;Residues: 1-868 <VAW>
A;Cross-references: EMBL:X77291; NID:g602888; PIDN:CAA54503.1; PID:g602896
R;van Dyck, L.; Pearce, D.A.; Sherman, F.
J. Biol. Chem. 269, 238-242, 1994
A;Title: PIM1 encodes a mitochondrial ATP-dependent protease that is required for mit
A;Reference number: S43937; MUID:94103216; PMID:8276800
A;Accession: S43939
A;Molecule type: DNA
A;Residues: 1-85 <VAN>
A;Cross-references: EMBL:X74544; NID:g453234; PIDN:CAA52635.1; PID:g453237
R;Tye, B.K.
submitted to GenBank, June 1990
A;Reference number: A39632
A;Accession: B39630
A;Molecule type: DNA
A;Residues: 1-163,'T',165-171,'IH',174-528,'P',530-577,'R',579-582,'H',584-711,'Q',71
A;Cross-references: GB:X53539; NID:g3911; PIDN:CAA37515.1; PID:g3912
R;Yan, H.; Gibson, S.; Tye, B.K.
Genes Dev. 5, 944-957, 1991
A;Title: Mcm2 and Mcm3, two proteins important for ARS activity, are related in struc
A;Reference number: A39630; MUID:91257581; PMID:2044961
A;Accession: A39630
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-79,'Y',81-163,'T',165-171,'IH',174-319,'R',321-528,'P',530-577,'R',579-
```

```
A;Cross-references: GB:X53539
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: SGD:MCM2; MIPS:YBL023c
A;Cross-references: SGD:S0000119; MIPS:YBL023c
A;Map position: 2L
C;Function:
A;Description: MCM2 is a component of the replication licensing factor that permits DNA
C;Superfamily: replication licensing factor MCM2; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;306-840/Domain: MCM homology <MCM>
F;341-367/Region: zinc finger CCCC motif

  Query Match             39.1%;  Score 1812;  DB 1;  Length 868;
  Best Local Similarity   46.3%;  Pred. No. 7.3e-107;
  Matches  384;  Conservative  112;  Mismatches  222;  Indels  112;  Gaps    9;

Qy          3 SSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEE 62
              |||  |  ||     |||            ||    |  :|     :     :||: :  :
Db         23 SSPQQHFRGGMNPVSSP----------IGSPDMINPEGDDNEVDDVPDIDEVEEQMNEVD 72

Qy         63 LIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACA 122
              |: |  |  || |    | |: :   :|| : :||: | |   |
Db         73 LMDDNMYEDYAADHNRDRYDPD--QVDDREQQELSLSERRRID---------------- 113

Qy        123 VGSCMTAMRRTRSALPASAASGAGTEDGEEDEQ--------------------MIESIEN 162
                 |   |  |  : |      | |::||                       | :||
Db        114 ------AQLNERDRLLRNVA----YIDDEDEEQEGAAQLDEMGLPVQRRRRRRQYEDLEN 163

Qy        163 LED----------------------LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHG 200
               :|                     :| :|  ||::       |    |:|| : |  |
Db        164 SDDDLLSDMDIDPLREELTLESLSNVKANSYSEWITQPNVSRTIARELKSFLLEYTDETG 223

Qy        201 HNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMY 260
               :|:  ||  : : | ||| |||  ||  : :|| || : | |:|:|||  |:|     |
Db        224 RSVYGARIRTLGEMNSESLEVNYRHLAESKAILALFLAKCPEEMLKIFDLVAMEATELHY 283

Qy        261 PKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNK 320
              | | || : ||||||  | :  || || :|: |:| :|||  ||| |||  ||:|| |
Db        284 PDYARIHSEIHVRISDFPTIYSLRELRESNLSSLVRVTGVVTRRTGVFPQLKYVKFNCLK 343

Qy        321 CNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLP 380
              |  :|||| |  |:|::    |  |:| ||| || |:|:|:|||| :||:|| |   |||
Db        344 CGSILGPFFQDSNEEIRISFCTNCKSKGPFRVNGEKTVYRNYQRVTLQEAPGTVPPGRLP 403

Qy        381 RSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKV 440
              | :: ||||||||    |:|:|:|||| ||||||:||  ||||||||:| || : :::
Db        404 RHREVILLADLVDVSKPGEEVEVTGIYKNNYDGNLNAKNGFPVFATIIEANSIKRREGNT 463

Qy        441 A--------VGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
              |        |   |:|: :    :|:|: | :|| :|:|||||| |||   | :||||
Db        464 ANEGEEEGLDVFSWTEEEEREFRKISRDRGIIDKIISSMAPSIYGHRDIKTAVACSLFGGV 523

Qy        493 PKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRH 552
              |||  ||| :|||||||| |||||||| |||:|| : ||:| |||||||| :|| |::
Db        524 PKNVNGKHSIRGDINVLLLGDPGTAKSQILKYVEKTAHRAVFATGQGASAVGLTASVRKD 583

Qy        553 PVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQAR 612
              |:::|||| |||||||:||||||||||||||||||||||||||||||||:||||
Db        584 PITKEWTLEGGALVLADKGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTTLQAR 643

Qy        613 CTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSH 672
              |::|||||| |||| :|  ::|| |||||:|||||||||| ||   || || ||| ||
Db        644 CSIIAAANPNGGRYNSTLPLAQNVSLTEPILSRFDILCVVRDLVDEEADERLATFVVDSH 703

Qy        673 VRHHPSNKE----EEGLANGSAA------------------EPAMPNTYGVEPLPQEVL 709
              || || | |    ||   || :|                  :       : |:|||:|
Db        704 VRSHPENDEDREGEELKNNGESAIEQGEDEINEQLNARQRRLQRQRKKEEEISPIPQELL 763

Qy        710 KKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESM 759
               ||| ||: :::|||:||| |||:::|:|||:||::||| ||||||:||:
Db        764 MKYIHYARTKIYPKLHQMDMDKVSRVYADLRRESISTGSFPITVRHLESI 813
```

```
RESULT 8
H90226
minichromosome maintenance protein MCM [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: H90226
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90226
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-686 <KUR>
A;Cross-references: GB:AE006641; NID:g13813947; PIDN:AAK41071.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO0774

  Query Match             19.7%;  Score 913;  DB 2;  Length 686;
  Best Local Similarity   36.2%;  Pred. No. 5.6e-50;
  Matches  216;  Conservative  115;  Mismatches  215;  Indels   50;  Gaps   11;

Qy        180 PRLEIHHR--FKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFL 237
              |  :| :|  |  || |   ::  | : |||:::   ::||:: : |: :    ||| :
Db          4 PSKQIDYRDVFIEFLTTFKGNNNQNKYIERINELVAYRKKSLIIEFSDVLSFNENLAYEI 63

Qy        238 PEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRT 297
                    :| | : |  : :| : | | |    :||||  :| | ||| :|   : :||
Db         64 INNTKIILPILEGALYDHILQLDPTYQRDIEKVHVRIVGIPRVIELRKIRSTDIGKLITI 123

Qy        298 SGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQS-----QNQEV-----KPGSCPECQSA 347
               |::   | |  ::    |        : | |       :::|:      |  ||:|
Db        124 DGILVKVTPVKERIYKATYK------HIHPDCMQEFEWPEDEEMPEVLEMPTICPKCGKP 177

Qy        348 GPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIY 407
              | | :  |:|   ::|:  ||| | :| : :|||  : ||  |||||   || :::|||
Db        178 GQFRLIPEKTKLIDWQKAVIQERPEEVPSGQLPRQLEIILEDDLVDSARPGDRVKVTGIL 237

Qy        408 HNNYDGSLNTANG--FPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 465
                  |  :   :  | ::  |       |  ::| : |   || | |  |:||  | ::
Db        238 DIKQDSPVKRGSRAVFDIYMKVSSIEVSQKVLDEVIISE---EDEKKIKDLAKDPWIRDR 294

Qy        466 IFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 525
              | :|||||||| ::|   |||||| ||      ::||||::|: |||||||| |::|
Db        295 IISSIAPSIYGHWELKEALALALFGGVPKVLEDT-RIRGDIHILIIGDPGTAKSQMLQFI 353

Qy        526 EKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 585
              :|: ||::|||:|:::|  :|| | |   : |: |||||||||| |: :||| ||| |:|
Db        354 SRVAPRAVYTTGKGSTAAGLTAAVVREKGTGEYYLEAGALVLADGGIAVIDEIDKMRDED 413

Qy        586 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 645
              | :|||||||::||:|||||  | ||  |||| ||  |||      |:|::|   |:||
Db        414 RVAIHEAMEQQTVSIAKAGIVAKLNARAAVIAAGNPKFGRYISERPVSDNINLPPTILSR 473

Qy        646 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 705
              ||:: :::|     ||  || ::: |     | |  : :                   :
Db        474 FDLIFILKDQPGE-QDRELANYILDVH-----SGKSTKNI----------------ID 509

Qy        706 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIP--ITVRHIESM 759
               : |:||| ||:: | ||:    :: :   : ::||:|  |   |  || | :|::
Db        510 IDTLRKYIAYARKYVTPKITSEAKNLITDFFVEMRKKSSETPDSPILITPRQLEAL 565


RESULT 9
B72775
probable DNA replication licensing factor mcm APE0188 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: B72775
```

```
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: B72775
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-699 <KAW>
A;Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79100.1; PID:g5103579
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0188
C;Superfamily: replication licensing factor MCM7; MCM homology

  Query Match             19.6%;  Score 909.5;  DB 2;  Length 699;
  Best Local Similarity   35.7%;  Pred. No. 9.5e-50;
  Matches  212;  Conservative  117;  Mismatches  216;  Indels   49;  Gaps   10;

Qy        182 LEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAP 241
              |  :  |||  ||        |   : | |  |      || | ::||   : :|:  | | |
Db         14 LAVGERFKTFLENFRTEEGKLKYVEAIRRMINYEETSLEVEFKDLYRYDPLLSEILLEKP 73

Qy        242 AELLQIFDEAALEVVLAMYPKY--DRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSG 299
              | |:   ||  |:|    |:|   |:      | : |   | :| :   |: :|:: :|
Db         74 REFLKEASEALKEIVAQESPEYAQGRV---FTPRFTGLFDTERIRDIGSDHVGKLVQING 130

Qy        300 VVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQE-------VKPGSCPEC-QSAGPFE 351
              :||       ::   ::  ::|      |    |::          :|  || | :  | |
Db        131 IVTRMHPRATRMVRARFRHDRCG---AEFWWPANEDEVLGERIERPSICPVCGEGGGKFT 187

Qy        352 VNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNY 411
              :  ::::| ::|:| :|| |  |   ::||| :  |  |||:    || ::: |:
Db        188 LVRDKSLYIDWQKIMVQERPEDVPGGQIPRSIEVHLSRDLVEKVRPGDRVKIVGVV---- 243

Qy        412 DGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIA 471
              |  : ::   :::  : || :  ::  :    :| || : |  ||:|  | ||| ||||
Db        244 -GLQSFSSSSTLYSLYMEANSILLEEKILEEVSITREDEEKILQLSRDPWIKEKIIASIA 302

Qy        472 PSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSR 531
              |:|||| |:|   || |||| ||      : ||||:||  |||| |||| |:   :|: |
Db        303 PTIYGHWDLKEAIALLLFGGVPKQRPDGTRTRGDIHVLFVGDPGVAKSQLLQSTAQVAPR 362

Qy        532 AIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHE 591
              ::|||:|::|  :|| | | | : |: ||||||||||: |: :|||||||: :|| |||
Db        363 VVYTTGKGSTAAGLTAAVLRDPRTGEYFLEAGALVLADGGIAVIDEFDKMSKEDRGVIHE 422

Qy        592 AMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCV 651
              |||||::||:||||  :| || :::|| ||  | |||| :| :|||| |||||||:: |
Db        423 AMEQQTVSIAKAGIKATLSARASLLAAGNPKFGYYDPSRSFVDNVDLPAPIISRFDLIFV 482

Qy        652 VRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKK 711
              ||| ::  :|||||  :|: :|              |    :|       ::|   ::|:|
Db        483 VRDVIERSRDEMLASYVLETH------------TNVELFKPE------IDP---DLLRK 520

Qy        712 YIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGS------IPITVRHIESM 759
              || :|:: | |:|    :  :   | ::|  ::   |      :||| | :|::
Db        521 YIAFARKHVKPRLTPQAKKLLKDFYVEMRSSALHHSSQEGAKPVPITTRQLEAL 574


RESULT 10
S65954
replication licensing factor MCM4 - human
N;Alternate names: cell division cycle control protein CDC21/CDC54; S. cerevisiae minich
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jan-2000
C;Accession: S65954; S43198; S41622
R;Musahl, C.; Schulte, D.; Burkhart, R.; Knippers, R.
Eur. J. Biochem. 230, 1096-1101, 1995
A;Title: A human homologue of the yeast replication protein Cdc21. Interactions with oth
A;Reference number: S65954; MUID:95324568; PMID:7601140
A;Accession: S65954
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'IRHERVATSSASGRLPRNARRRGRALARSRRERPRCGAGQGSREAGPARACPCRAGTPST',1-863 <MUS
A;Cross-references: EMBL:X74794; NID:g683749; PIDN:CAA52801.1; PID:g940536
R;Hu, B.
submitted to the EMBL Data Library, August 1993
A;Reference number: S43198
A;Accession: S43198
A;Molecule type: mRNA
A;Residues: 'ARE',440-702 <HUB1>
A;Cross-references: EMBL:X74794
R;Hu, B.; Burkhart, R.; Schulte, D.; Musahl, C.; Knippers, R.
Nucleic Acids Res. 21, 5289-5293, 1993
A;Title: The P1 family: a new class of nuclear mammalian proteins related to the yeas
A;Reference number: S41622; MUID:94089373; PMID:8265339
A;Accession: S41622
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'ARE',440-636 <HUB2>
A;Cross-references: EMBL:X74794
C;Comment: The complex of six MCM proteins is one of several proteins that must be bo
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: GDB:MCM4; CDC21; CDC54
A;Cross-references: GDB:433798
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, a
C;Function:
A;Description: part of the replication licensing system that permits DNA replication
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; p
F;271-764/Domain: MCM homology <MCM>

  Query Match             18.9%;  Score 877;  DB 1;  Length 863;
  Best Local Similarity   30.4%;  Pred. No. 1.5e-47;
  Matches  246;  Conservative  129;  Mismatches  280;  Indels  154;  Gaps   23;

Qy          3 SSPAQRRRGND--------PLTSSPG---RSSRRTDALTSSPGRDLPPFEDESEGLLGTE 51
              |||:|||||  |        |: :|||   :|:   | | |||
Db         31 SSPSQRRRGEDSTSTGELQPMPTSPGVDLQSTAAQDVLFSSP------------------ 72

Qy         52 GPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT----ASRREAA--D 105
                                        |::    :  : || :    ||    :|| |
Db         73 ------------------------PQM---HSSAIPLDFDVSSPLTYGTPSSRVEGTPRS 105

Qy        106 GPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
              |  ||     | ||:   |  :               : ||  ||    || :
Db        106 GVRGTPVRQRPDLGSAQKGLQVDLQ----------SDGAAAEDIVASEQSL-------- 146

Qy        166 LKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVD---SHGHNV--------FKERISDM 211
                |  :  |   |::|  :      |: ||:  :|      ||        : :|: ::
Db        147 --GQKLVIWGTDVNVAACK----ENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEI 200

Qy        212 CKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNH-I 270
                  : | || | : : :  |   |   | |::  || |  |:    ||  | |  | |
Db        201 NVIGEQFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYP--DSILEHQI 258

Qy        271 HVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 330
              ||  :    : :|:|    ::|||  ||:|   : ::|::    : |  |
Db        259 QVRPFNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTT-RVEM 317

Qy        331 SQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLA- 389
              : :  :|  |  | :    :    ::: : | |::||||  : | : | :   || |
Db        318 DRGRIAEPSVCGRCHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHT--VILFAH 375

Qy        390 -DLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATVILANHVAKKDNKVAV 442
              ||||    || : :||||       |        :|   |: | |   |  | | |
Db        376 NDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVSNVKSVYKTHIDVIHYRKTDAKRLH 430

Qy        443 GELTDED----------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
              |   ||:          |:::  ||:   | |:: :::||||| |||||:|  | ||||
Db        431 G--LDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGT 488

Qy        493 PK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQ 550
              |  :  |: | | :||:|||||||:||| |:|:  :  |  :|:|:|:||| :||||
```

```
Db        489 RKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVM 548

Qy        551 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 610
              : | :|:  |: |||||:| |:| |||||||:  |: :|| ||||::||:||||:  |
Db        549 KDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLN 608

Qy        611 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVG 670
              || :|:|||||  :::|  |  ||: |   ::||||:: :: |  |   |  ||  :|
Db        609 ARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVA 668

Qy        671 SHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD 730
              : :    :|                    | |   ||| || ||   : |:|::
Db        669 LYYQSEEQAEE-------------------ELLDMAVLKDYIAYAHSTIMPRLSEEASQ 708

Qy        731 KVAKMYSDLRKESMATGSIPITVRHIESM 759
              : : | |:||   : | :    | :||:
Db        709 ALIEAYVDMRKIGSSRGMVSAYPRQLESL 737


RESULT 11
S64720
replication licensing factor MCM4 - African clawed frog
N;Alternate names: cell division cycle control protein CDC21/CDC54
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C;Accession: S64720; S26643; S25529
R;Coue, M.; Kearsey, S.E.; Mechali, M.
EMBO J. 15, 1085-1097, 1996
A;Title: Chromatin binding, nuclear localization and phosphorylation of Xenopus cdc21 ar
A;Reference number: S64720; MUID:96183193; PMID:8605878
A;Accession: S64720
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-863 <COU>
A;Cross-references: EMBL:U29178; NID:g1002597; PIDN:AAB01680.1; PID:g1002598
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
R;Coxon, A.; Maundrell, K.; Kearsey, S.E.
Nucleic Acids Res. 20, 5571-5577, 1992
A;Title: Fission yeast cdc21(+) belongs to a family of proteins involved in an early ste
A;Reference number: S26640; MUID:93087163; PMID:1454522
A;Accession: S26643
A;Molecule type: DNA
A;Residues: 513-523,'Y',525-538,'G',540-553,'G',555-588 <COX>
A;Cross-references: EMBL:Z15033; NID:g64612; PIDN:CAA78751.1; PID:g64613
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Genetics:
A;Gene: cdc21
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;271-764/Domain: MCM homology <MCM>

  Query Match             18.9%;  Score 874.5;  DB 1;  Length 863;
  Best Local Similarity   28.1%;  Pred. No. 2.1e-47;
  Matches  264;  Conservative  145;  Mismatches  358;  Indels  173;  Gaps   25;

Qy          3 SSPAQRR----RGNDPLT------SSPGRSSRRTDALTS---------SPGRDLPPFEDE 43
              |:|::||    ||::| |       ||    |||:  ||         ||  |:
Db          6 STPSRRRNKRGRGSNPPTPHGEEVQSPPSQRRRTEDSTSIGELLPMPTSPSGDV------ 59

Qy         44 SEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREA 103
                    :   |:||:          :|   :     | |          :|| |
Db         60 -----------QSPSGQELLFSS------PVPSRHSAHQSELDLSSPLTYGTPSSRVEG 101

Qy        104 A--DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIE 161
                 |  ||     | ||:          |: :               |   |:   | :
Db        102 TPRSGIRGTPARQRPDLGSA---------RKVKQV-----------DLHSDQPAAEELV 140

Qy        162 NLEDLKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVDSHGHN----------VFKER 207
                  |   |  :  |   |::|        :|: |::  :|               :: :|

Db        141 TSEQSLGQKLVIWGTDVNVA----TCKEKFQRFVQRFIDPSAKEEDNVGLDLNEPIYMQR 196

Qy        208 ISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRIT 267
              : ::      | :: : |   :  |   |  | |::  || || |:    ||  | |
Db        197 LEEINVVGDPFLNIDCDHLRNFDQDLYRQLVCYPQEVIPTFDMAANEIFFERYP--DSIL 254

Qy        268 NH-IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLG 326
               | | ||  :      :|||    ::|||  ||:|   : ::|::    : |  | |
Db        255 EHQIQVRPYNALKTRNMRSLNPEDIDQLITISGMVIRTSQIIPEMQEAFFKCQVCAFTTR 314

Qy        327 PFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAI 386
                  : :  :|  |  | :     :    ::: : | |::||||  : | : | :
Db        315 VEI-DRGRIAEPSVCKHCNTTHSMALIHNRSMFSDKQMIKLQESPEDMPAGQTPHTTILY 373

Qy        387 LLADLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATVILANHVAKKDNKV 440
                ||||    || : :||||      |          |   |: | |   |  | |:|
Db        374 GHNDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVRNVKSVYKTHIDVIHYRKTDSKR 428

Qy        441 AVG--------ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
                |          |:| | |:  |:    | |:: |::||||| |||||:|  | ||||
Db        429 LHGIDEDTEQKLFTEERVAMLKELAAKPDIYERLAAALAPSIYEHEDIKKGILLQLFGGT 488

Qy        493 PK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQ 550
              |  :  |: | | ::|:|||||||:||| |:|:  :  |  :|:|:|:||| :||||
Db        489 RKDFSHTGRGKFRAEVNILLCGDPGTSKSQLLQYVFNLVPRGQYTSGKGSSAVGLTAYVM 548

Qy        551 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 610
              : | :|:  |: |||||:| |:| |||||||:  |: :|| ||||::||:||||:  |
Db        549 KDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLN 608

Qy        611 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVG 670
              || :|:|||||:  :::|  |  ||: |   ::||||:: :: |  |   |  ||  :|
Db        609 ARTSVLAAANPVESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVA 668

Qy        671 SHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD 730
              : :    |||                    |   ||| || ||:  |:|:|::
Db        669 LYYQSEEQMKEEH-------------------LDMAVLKDYIAYARTYVNPRLSEEASQ 708

Qy        731 KVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPR----- 785
              : : |  :||     | :    | :||:                ||   |   |
Db        709 ALIEAYVSMRKIGSGRGMVSAYPRQLESLI--------------RRAEAHAKVRFSNKV 753

Qy        786 ------DAGELHRHTEVQRHRSMRK--------TFARYLSFRRDNNELLLFILKQLVAEQ 831
                   :|  |||     |  :     |      :  |:   ||   :||:|:  :
Db        754 ETIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATARKRKEELAQ-VLKKLIQSK 812

Qy        832 -----VTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLS 866
                  : ||:     :  :     ||: |:|    :| |:
Db        813 GKTPALKYQQLFEDLRGQSDAAITKDMFDEA----LHALA 848


RESULT 12
T41059
replication licensing factor MCM3 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: minichromosome maintenance protein MCM3
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C;Accession: T41059; T43795
R;Murphy, L.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21968
A;Accession: T41059
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-879 <MUR>
A;Cross-references: EMBL:AL031525; PIDN:CAA20668.1; GSPDB:GN00068; SPDB:SPCC1682.02c
A;Experimental source: strain 972h-; cosmid c1682
R;Sherman, D.A.; Forsburg, S.L.
Nucleic Acids Res. 26, 3955-3961, 1998
A;Title: Schizosaccharomyces pombe Mcm3p, an essential nuclear protein, associates ti
A;Reference number: Z22684; MUID:98371232; PMID:9705504
A;Accession: T43795
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
A;Molecule type: DNA
A;Residues: 1-879 <SHE>
A;Cross-references: EMBL:AF063864; NID:g3139136; PIDN:AAC32263.1; PID:g3139137
A;Experimental source: strain Sp.011
C;Genetics:
A;Gene: mcm3; SPCC1682.02c
A;Map position: 3
C;Superfamily: human replication licensing factor MCM3; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; nucleus; phosphoprotein
F;117-665/Domain: MCM homology <MCM>

  Query Match             18.9%;  Score 873;  DB 2;  Length 879;
  Best Local Similarity   34.1%;  Pred. No. 2.7e-47;
  Matches  219;  Conservative  108;  Mismatches  225;  Indels   90;  Gaps   13;

Qy        188 FKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQI 247
Db         17 FQEYLEHDTDDANVTLYQEAILRMLNMGQRRLIVNIDELRDYNRELADGVLLKPLEYVEP 76

Qy        248 FDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEEL---------------RSLRQLHL 291
Db         77 FDEALRNVVSTLIDP-----------VVHKDLKDKLFYVGFRGSFGDHHVNPRTLRAMHL 125

Qy        292 NQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC-------PEC 344
Db        126 NKMISLEGIVTRCSFVRPKVIKSVHYCEATK--RHHFKQYADATMNGGLSFQSTVYPTQD 183

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
Db        184 ENGNPLSIEFGFSTFRDHQSISLQEMPERAPPGQLPRSIDILLDDDLVDTVKPGDRVNIV 243

Qy        405 GIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVG-------ELTDEDVKMITSLS 457
Db        244 GQYRSM--GSKTSGNTSATFRTVLLANNVVLLGNKPGLGNVGGGALDITDADIRNINKLA 301

Qy        458 KDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKN-PGGKHKVRGDINVLLCGDPGT 516
Db        302 RKKNVFELLSTSLAPSIYGYEYVKQAILLLLLGGTEKNLTNGTH-IRGDINILMVGDPST 360

Qy        517 AKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLID 576
Db        361 AKSQLLRFVLNTAPLAIATTGRGSSGVGLTAAVTTDKETGERRLEAGAMVLADRGVVCID 420

Qy        577 EFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENV 636
Db        421 EFDKMSDIDRVAIHEVMEQQTVTIAKAGIHTSLNARCSVIAAANPIYGQYDIRKDPHQNI 480

Qy        637 DLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEE--------EGLANG 688
Db        481 ALPDSMLSRFDLLFIVTDDIDDKKDRALSEHVLRMH-RYLPPGVEPGTPVRDSLNSVLNV 539

Qy        689 SAAEPAMPNTYGVEPLPQEV---------------------------LKKYIIYAKERV 720
Db        540 GATNAAGVSTENVE---QEVETPVWETFSSLLHANARTKKKELLNINFVRKYIQYAKSRI 596

Qy        721 HPKLNQMDQDKVAKMYSDLRKESMATG---SIPITVRHIESM 759
Db        597 HPILNQATAEYITNIYCGLRNDDLQGNQRRTSPLTARTLETL 638
```

RESULT 13
C84540
probable CDC21 protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84540
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84540

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-720 <STO>
A;Cross-references: GB:AE002093; NID:g4544386; PIDN:AAD22296.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g16440
A;Map position: 2

  Query Match             18.5%;  Score 858;  DB 2;  Length 720;
  Best Local Similarity   29.2%;  Pred. No. 1.8e-46;
  Matches  238;  Conservative  132;  Mismatches  264;  Indels  182;  Gaps   22;

Qy        133 TRSALPASAASGAGTEDGEED--------------------------EQMIESIENLE 164
Db          3 TDEPLPSS-------DDGEEDGGDDTTPTFVWGTNISVQDVKSAIEMFVKHFREARENSD 55

Qy        165 DLKGHSVREWVSMAGPR--LEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVN 222
Db         56 DL----FREGKYMVSIRKVIEIEGEW-------IDVDAFDVF----------------D 87

Qy        223 YE-DLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 281
Db         88 YDPDLYNK-------MVRYPLEVLAIFDIVLMDIVSTINRLFEK---HVQVRIFNLRTST 137

Qy        282 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC 341
Db        138 SMRNLNPSDIEKMISLKGMIIRSSSIIPEIREAVFRCLVCGYFSDPIIVDRGKISEPPTC 197

Qy        342 --PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGD 399
Db        198 LKQECMTKNSMTLVHNRCRFADKQIVRLQETPDEIPEGGTPHTVSLLLHDKLVDNGKPGD 257

Qy        400 EIELTGIYHNNYDGSLNTANGFP-------VFATVILANHVAKKDNKVAVG--------- 443
Db        258 RIEVTGIYR------AMTVRVGPAHRTVKSVFKTYIDCLHI-KKASKLRMSAEDPMDVDN 310

Qy        444 --ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHK 501
Db        311 SLRRVDEDLRKFQELSKQPDIYERLSRSLAPNIWELDDVKKGLLCQLFGGNALNLASGAN 370

Qy        502 VRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLE 561
Db        371 FRGDINILLVGDPGTSKSQLLQYIHKLSPRGIYTSGRGSSAVGLTAYVAKDPETGETVLE 430

Qy        562 AGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANP 621
Db        431 SGALVLSDRGICCIDEFDKMSDSARSMLHEVMEQQTVSIAKAGIIASLNARTSVLACANP 490

Qy        622 IGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE 681
Db        491 SGSRYNPRLSVIENIHLPPTLLSRFDLIYLILDKPDEQTDRRLAKHIVALHFENAESAQE 550

Qy        682 EEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK 741
Db        551 -------------------EAIDITTLTTYVSYARKNIHPKLSDEAAEELTRGYVELRK 590

Qy        742 ESMATGS----IPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGELHR----- 792
Db        591 AGKFAGSSKKVITATPRQIESL-------IRLSEALARMRFSEWVEKHDVDEAFRLLRVA 643

Qy        793 -----------------HTEVQRHRSMRK-TFA---RYLSFRRDNNELLLFILKQLVAE 830
Db        644 MQQSATDHATGTIDMDLINTGVSASERMRRDTFASSIRDIALEKITMEMLLL-------- 695

Qy        831 QVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLS 866
Db        696 QLLEELKKHGGNINT------------EIHLHDVS 718
```

RESULT 14
S56766
replication licensing factor MCM4 - mouse
N;Alternate names: cell division cycle control protein CDC21/CDC54

```
C;Species: Mus musculus (house mouse)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 16-Jun-2000
C;Accession: S56766
R;Kimura, H.; Takizawa, N.; Nozaki, N.; Sugimoto, K.
Nucleic Acids Res. 23, 2097-2104, 1995
A;Title: Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs and character
A;Reference number: S56766; MUID:95334361; PMID:7610039
A;Accession: S56766
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-862 <KIM>
A;Cross-references: EMBL:D26089; NID:g940405; PIDN:BAA05082.1; PID:g940406
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
C;Comment: The complex of six MCM proteins is one of several proteins that must be bound
 phosphorylated and dissociate from the chromatin.
C;Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, and
C;Function:
A;Description: part of the replication licensing system that permits DNA replication to
C;Superfamily: replication licensing factor MCM4; MCM homology
C;Keywords: cell cycle control; DNA replication initiation; heterohexamer; nucleus; phos
F;270-763/Domain: MCM homology <MCM>

  Query Match             18.5%;  Score 856;  DB 1;  Length 862;
  Best Local Similarity   29.0%;  Pred. No. 3.2e-46;
  Matches  275;  Conservative  140;  Mismatches  331;  Indels  202;  Gaps   33;

Qy          2 ASSPAQR--RRG---------NDPLTSSPGRSSRRTDALT-------SSPGRDL--PPFE 41
              ||:|::|  |||         ::   ||| |  |  |: |       :||| ||  || :
Db          5 ASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSSTGELLPMPTSPGADLQSPPAQ 64

Qy         42 DESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT---- 97
              :          |     :    |||           || :    ||
Db         65 N------------------ALFSSPPQMHSLAIP-----------LDFDVSSPLTYGTP 94

Qy         98 ASRREAA--DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQ 155
              :|| |     |  ||     | ||:             |  |     |     ||   |
Db         95 SSRVEGTPRSGVRGTPVRQRPDLGS-----------ARKGLQVDLQS-----DGAAAED 137

Qy        156 MIESIENLEDLKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVD---SHGHNV------ 203
              :: | ::|    |  :  |   |::|         |: ||:   |       ||
Db        138 IVPSEQSL----GQKLVIWGTDVNVA----TCKENFQRFLQCFTDPLAKEEENVGIDITQ 189

Qy        204 --FKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYP 261
                : ::: ::       | || | : :    |   |   | |::  || |  |:    ||
Db        190 PLYMQQLGEINITGEPFLNVNCEHIKSFSKNLYRQLISYPQEVIPTFDMAVNEIFFDRYP 249

Qy        262 KYDRITNH-IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNK 320
              | |  | | ||  :    : :|:|    ::|||  ||:|   : ::|::    : |
Db        250 --DSILEHQIQVRPFNALKTKSMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQV 307

Qy        321 CNFVLGPFCQSQNQEV------KPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKV 374
              |        :   |:      :| ||  | :     :    : : : | |::|||  :
Db        308 C-------AHTTRVEIDRGRIAEPCSCVHCHTTHSMALIHNRSFFSDKQMIKLQESPEDM 360

Qy        375 AARRLPRSKDAILLA--DLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFAT 426
              | : | :   :| |  ||||    || : :||||       |       :|   |: |
Db        361 PAGQTPHT--IVLFAHNDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVSNVKSVYKT 413

Qy        427 VILANHVAKKDNKVAVGELTDED----------VKMITSLSKDQQIGEKIFASIAPSIYG 476
              |   |  | | |   |   ||:          ||::  ||:   | |:: :::||||
Db        414 HIDVIHYRKTDAKRLHG--LDEEAEQKLFSEKRVKLLKELSRKPDIYERLASALAPSIYE 471

Qy        477 HEDIKRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIF 534
              |||||:|  | ||||  |  :   |: | | :||:||||||||:||| |:|:  :  |  :
Db        472 HEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQY 531

Qy        535 TTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAME 594
              |:|:|:||| :|||| : | :|:  |: ||||||:| |:| ||||||||:  |: :|| ||
Db        532 TSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVME 591

Qy        595 QQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRD 654
              ||::||:||||:  | || :|:|||||  :::|  |  ||: |   ::||||:: :: |
Db        592 QQTLSIAKAGIICQLNARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLD 651

Qy        655 TVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYII 714
              |   |  ||  :|  : :     :||                       |   ||| ||
Db        652 PQDEAYDRRLAHHLVSLYYQSEEQVEEEF-------------------LDMAVLKDYIA 691

Qy        715 YAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLC 774
              ||   : |:|::     : : | ::||   : | :    | :||:
Db        692 YAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLI-------------- 737

Qy        775 DRRRRQHGHPR----------DAGELHRHTEVQRHRSMRK--------TFARYLSFRRD 815
              |   |  |             :|  |||    |     |         |     : |:
Db        738 -RLAEAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKR 796

Qy        816 NNELLLFILKQLVAEQVT----YQR--NRFGAQQDTIEVPEKDLVDKA 857
              ||   : | ::::  |    ||:       | ||     ||: ::|
Db        797 KEELAEALRKLILSKGKTPALKYQQLFEDIRGQSDT--AITKDMFEEA 842


RESULT 15
T52118
probable replication licensing factor MCM3 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 07-Mar-2003
C;Accession: T52118
R;Dodeman, V.L.; Phan, T.; Sabelli, P.; Bergounioux, C.
submitted to the EMBL Data Library, September 1997
A;Description: Expression analysis of Arabidopsis thaliana MCM3 homolog during the ce
A;Reference number: Z25964
A;Accession: T52118
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-776 <DOD>
A;Cross-references: EMBL:AJ000058; PIDN:CAA03887.1
C;Superfamily: human replication licensing factor MCM3; MCM homology

  Query Match             18.5%;  Score 855;  DB 2;  Length 776;
  Best Local Similarity   32.6%;  Pred. No. 3.2e-46;
  Matches  232;  Conservative  117;  Mismatches  283;  Indels   80;  Gaps   15;

Qy        181 RLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEA 240
              ||     |  ||          ::: | |  :  : |  |::|  |:      :|  : :
Db          8 RLRHKRDFIQFL--------DSMYMEEIKALVHQKRHRLIINISDIHHHFREVASRILKN 59

Qy        241 PAELLQIFDEAALEVVLAMYPKYDRITNHIHVRIS--HLPLVEELRSLRQLHLNQLIRTS 298
              | | :| | :|| |   |: ||| :    : |      :  |   | |    :  ::
Db         60 PNEYMQSFCDAATEATRAIDPKYLKEGELVLVGFEGYFVSRVVTPRELLSDFIGSMVCVE 119

Qy        299 GVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETI 358
              |:|| |: | |::    : |       | |     :::   :     |  |    :   :
Db        120 GIVTKCSLVRPKVVKSVHFCPS----TGEFTNRDYRDITSHAGLPTGSVYPTRDDKGNLL 175

Qy        359 --------YQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNN 410
                     |:::| : ||| |   |  :|||| | |   ||||||  || : : |||
Db        176 VTEYGLCKYKDHQTLSIQEVPENAAPGQLPRSVDVIAEDDLVDSCKPGDRVSVFGIYKAL 235

Qy        411 YDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASI 470
                |  : ||  || |:::||::|  : :      | :|:  | ::::     : :  |:
Db        236 PGKSKGSVNG--VFRTILIANNIALLNKEANAPIYTKQDLDNIKNIARRDDAFDLLARSL 293

Qy        471 APSIYGHEDIKRGPALALFGGEPKN-PGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVS 529
              |||||||  ||:   | : ||  ||   | | :|||||:: |||  |||| |: |  ::
Db        294 APSIYGHAWIKKAVVLLMLGGVEKNLKNGTH-LRGDINMMMVGDPSVAKSQLLRAIMNIA 352

Qy        530 SRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSI 589
              || |||:|:| | :|| |     : |  |||||:||||:|:  ||||||||||| :|
Db        353 PLAISTTGRGSSGVGLTAAVTSDQETGERRLEAGAMVLADKGIVCIDEFDKMNDQDRVAI 412

Qy        590 HEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDIL 649
              || ||||:::|:||||  || |||:|:|||||| | || ||| ::|: | : ::|||||:|
Db        413 HEVMEQQTVTIAKAGIHASLNARCSVVAAANPIYGTYDRSLTPTKNIGLPDSLLSRFDLL 472

Qy        650 CVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEG------LANGSAAEPAMPNTYGV-- 701
              :| | :|   | |::  |:  |   : ::: | |       |    ||  |   |
```

```
Db        473 FIVLDQMDAGIDSMISEHVLRMH--RYKNDRSEAGPDGSLPYAREDNAESEMFVKYNQTL 530

Qy        702 ----------EPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMAT---G 747
                      : |   : ||||| ||| |: |||     :::|: |:|||     |   |
Db        531 HGKKKRGQTHDKTLTIKFLKKYIHYAKHRITPKLTDEASERIAEAYADLRNAGSDTKTGG 590

Qy        748 SIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHG----------HPRDAGELHRHTEV 796
              ::||| | :|::       ||    |     :            : ::  |:    :
Db        591 TLPITARTLETIIRLA--TAHAKMKLSSEVTKADAEAALKLMNFAIYHQELTEMDDREQE 648

Qy        797 QRHRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFGAQQDTIEV 848
              :| |   :   |  | || |                 : |  ||: ||  |
Db        649 ERQREQAEQ-ERTPSGRRGNQR---------------RNNEDGAENDTANV 683


Search completed: September 12, 2003, 13:14:22
Job time : 31 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 13:09:25 ; Search time 16 Seconds
(without alignments)
2621.739 Million cell updates/sec

Title: CAA47749
Perfect score: 4630
Sequence: 1 MASSPAQRRRGNDPLTSSPG.....LFRMNKFSHDLKRKMILQQF 892

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4078.5	88.1	895	1	MCM2_HUMAN	P49736 homo sapien
2	3941	85.1	904	1	MCM2_MOUSE	P97310 mus musculu
3	3539.5	76.4	886	1	MCM2_XENLA	P55861 xenopus lae
4	2649	57.2	887	1	MCM2_DROME	P49735 drosophila
5	1951.5	42.1	830	1	CC19_SCHPO	P40377 schizosacch
6	1812	39.1	868	1	MCM2_YEAST	P29469 saccharomyc
7	877	18.9	863	1	MCM4_HUMAN	P33991 homo sapien
8	874.5	18.9	863	1	MCM4_XENLA	P30664 xenopus lae
9	873	18.9	879	1	MCM3_SCHPO	P30666 schizosacch
10	856	18.5	862	1	MCM4_MOUSE	P49717 mus musculu
11	848.5	18.3	933	1	CC54_YEAST	P30665 saccharomyc
12	843	18.2	931	1	CC21_SCHPO	P29458 schizosacch
13	840	18.1	821	1	MCM6_MOUSE	P97311 mus musculu
14	835	18.0	821	1	MCM6_HUMAN	Q14566 homo sapien
15	825.5	17.8	735	1	MCM5_XENLA	P55862 xenopus lae
16	816.5	17.6	866	1	MCM4_DROME	Q26454 drosophila
17	802	17.3	971	1	MCM3_YEAST	P24279 saccharomyc
18	793.5	17.1	1017	1	MCM6_YEAST	P53091 saccharomyc
19	789.5	17.1	734	1	MCM5_HUMAN	P33992 homo sapien
20	789	17.0	775	1	MCM5_YEAST	P29496 saccharomyc
21	786.5	17.0	892	1	MIS5_SCHPO	P49731 schizosacch
22	783	16.9	720	1	MCM7_XENLA	Q91876 xenopus lae
23	781	16.9	719	1	MCM7_HUMAN	P33993 homo sapien
24	778.5	16.8	808	1	MCM3_HUMAN	P25205 homo sapien
25	777.5	16.8	716	1	PROL_ARATH	P43299 arabidopsis
26	776.5	16.8	719	1	MCM7_MOUSE	Q61881 mus musculu
27	772.5	16.7	720	1	NDA4_SCHPO	P41389 schizosacch
28	770	16.6	760	1	MCM7_SCHPO	O75001 schizosacch
29	769	16.6	807	1	MCM3_XENLA	P49739 xenopus lae
30	768	16.6	733	1	MCM5_MOUSE	P49718 mus musculu
31	764.5	16.5	812	1	MCM3_MOUSE	P25206 mus musculu
32	761	16.4	759	1	MCM5_CAEEL	Q21902 caenorhabdi
33	758.5	16.4	840	1	MCM8_HUMAN	Q9uja3 homo sapien
34	751	16.2	810	1	MCM6_CAEEL	P34647 caenorhabdi
35	748.5	16.2	600	1	MCM3_MAIZE	Q43704 zea mays (m
36	748	16.2	845	1	CC47_YEAST	P38132 saccharomyc
37	650	14.0	507	1	MCM6_RAT	Q62724 rattus norv
38	644	13.9	597	1	MCM3_ENTHI	Q24849 entamoeba h
39	591	12.8	495	1	MCM8_MOUSE	Q9cwv1 mus musculu
40	466.5	10.1	682	1	YE89_METJA	Q58884 methanococc
41	442	9.5	762	1	Y961_METJA	Q58371 methanococc
42	430	9.3	759	1	Y363_METJA	Q57809 methanococc
43	271	5.9	602	1	YZ13_METJA	Q60275 methanococc
44	154.5	3.3	516	1	YIFB_ECOLI	P22787 escherichia
45	137.5	3.0	506	1	YIFB_SALTY	P57015 salmonella

ALIGNMENTS

RESULT 1
MCM2_HUMAN
ID MCM2_HUMAN STANDARD; PRT; 895 AA.
AC P49736; Q14577; Q15023; Q969W7; Q96AE1; Q9BRM7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA replication licensing factor MCM2 (Nuclear protein BM28).
GN MCM2 OR CDCL1 OR BM28 OR KIAA0030.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94230605; PubMed=8175912;
RA Todorov I.T., Pepperkok R., Philipova R.N., Kearsey S.E., Ansorge W.,
RA Werner D.;
RT "A human nuclear protein with sequence homology to a family of early
RT S phase proteins is required for entry into S phase and for cell
RT division.";
RL J. Cell Sci. 107:253-265(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.";
RL DNA Res. 1:27-35(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Mimura S., Nishimoto S., Kubota Y., Takisawa H., Nojima H.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Lung, Lymph, Muscle, Placenta, and Uterus;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

```
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   MAPPING.
RX   MEDLINE=94080845; PubMed=8258304;
RA   Mincheva A., Todorov I., Werner D., Fink T.M., Lichter P.;
RT   "The human gene for nuclear protein BM28 (CDCL1), a new member of the
RT   early S-phase family of proteins, maps to chromosome band 3q21.";
RL   Cytogenet. Cell Genet. 65:276-277(1994).
CC   -!- FUNCTION: ACTS AS A FACTOR THAT ALLOWS THE DNA TO UNDERGO A SINGLE
CC       ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR THE ENTRY IN
CC       S PHASE AND FOR CELL DIVISION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67334; CAA47749.1; -.
DR   EMBL; D21063; BAA04642.1; ALT_INIT.
DR   EMBL; D83987; BAA12177.1; -.
DR   EMBL; BC006165; AAH06165.1; -.
DR   EMBL; BC007670; AAH07670.1; -.
DR   EMBL; BC007938; AAH07938.1; -.
DR   EMBL; BC017258; AAH17258.1; -.
DR   EMBL; BC014272; AAH14272.1; -.
DR   EMBL; BC017490; AAH17490.1; -.
DR   PIR; S42228; S42228.
DR   Genew; HGNC:6944; MCM2.
DR   GK; P49736; -.
DR   MIM; 116945; -.
DR   GO; GO:0005717; C:chromatin; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003677; F:DNA binding activity; TAS.
DR   GO; GO:0007049; P:cell cycle; TAS.
DR   GO; GO:0006260; P:DNA replication; TAS.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger; ATP-binding; Cell cycle.
FT   ZN_FING      320    346       C4-TYPE (POTENTIAL).
FT   DOMAIN       464    671       MCM.
FT   NP_BIND      514    521       ATP (POTENTIAL).
FT   CONFLICT     100    100       Q -> R (IN REF. 1).
FT   CONFLICT     105    117       ERAMRQRDREAGR -> DGPCGTVTGSWP (IN
FT                                 REF. 1).
FT   CONFLICT     121    148       RMRRGLLYDSDEEDEERPARKRRQVERA ->
FT                                 ACAVGSCMTAMRRTRSALPASAASGAG (IN REF. 1).
FT   CONFLICT     157    157       E -> Q (IN REF. 1).
FT   CONFLICT     379    379       G -> R (IN REF. 1).
FT   CONFLICT     398    399       KP -> NA (IN REF. 1 AND 3).
FT   CONFLICT     486    486       L -> P (IN REF. 1).
FT   CONFLICT     492    492       G -> R (IN REF. 4; AAH17258).
FT   CONFLICT     546    547       GL -> AV (IN REF. 1).
FT   CONFLICT     718    718       A -> T (IN REF. 4; AAH06165).
FT   CONFLICT     762    802       IRMAEAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVM
FT                                 -> SHGGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQ
FT                                 RH (IN REF. 1).
SQ   SEQUENCE   895 AA;  100925 MW;  61ACDAAAC77B3700 CRC64;

  Query Match             88.1%;  Score 4078.5;  DB 1;  Length 895;
  Best Local Similarity   89.8%;  Pred. No. 1.3e-249;

    Matches  817;  Conservative    8;  Mismatches   52;  Indels   33;  Gaps    6;

Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              |||||||||||||||||||||||||||||||||||||||||:||||       |   |
Db         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQRDREAGR--- 117

Qy        114 SWPGLGACAVGSCMTA--------MRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
                 |||    |    :         |: |    |       ||||||||:||||||||||
Db        118 ---GLGRMRRGLLYDSDEEDEERPARKRRQVERA-------TEDGEEDEEMIESIENLED 167

Qy        166 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 227

Qy        226 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 287

Qy        286 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 347

Qy        346 SAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTG 405
              |||||||||||||||||||||||||||||| ||||||||||||||||  ||||||||
Db        348 SAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTG 407

Qy        406 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 465
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 467

Qy        466 IFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 525
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        468 IFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 527

Qy        526 EKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 585
              |||||||||||||||| :||||||||||||||||||||||||||||||||||||||||||||
Db        528 EKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 587

Qy        586 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 647

Qy        646 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 707

Qy        706 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGP 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        708 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA-- 765

Qy        766 RAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
               ||    | |      :      |    : |:    |||||||||||||||||||||| |
Db        766 EAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLF 825

Qy        823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        826 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 885

Qy        883 LKRKMILQQF 892
              ||||||||||
Db        886 LKRKMILQQF 895


RESULT 2
MCM2_MOUSE
ID   MCM2_MOUSE     STANDARD;      PRT;   904 AA.
AC   P97310; O89057;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
```

```
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA replication licensing factor MCM2.
GN   MCM2 OR MCMD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 217-224; 376-384 AND 477-487.
RX   MEDLINE=97224213; PubMed=9077461;
RA   Kimura H., Ohtomo T., Yamaguchi M., Ishii A., Sugimoto K.;
RT   "Mouse MCM proteins: complex formation and transportation to the
RT   nucleus.";
RL   Genes Cells 1:977-993(1996).
RN   [2]
RP   REVISIONS.
RA   Kimura H.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 477-668 FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Spleen;
RX   MEDLINE=99012997; PubMed=9798653;
RA   Chu C.C., Paul W.E.;
RT   "Expressed genes in interleukin-4 treated B cells identified by cDNA
RT   representational difference analysis.";
RL   Mol. Immunol. 35:487-502(1998).
CC   -!- FUNCTION: ACTS AS A FACTOR THAT ALLOWS THE DNA TO UNDERGO A SINGLE
CC       ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR THE ENTRY IN
CC       S PHASE AND FOR CELL DIVISION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86725; BAA22148.1; -.
DR   EMBL; U89403; AAC36510.1; -.
DR   PIR; T10067; T10067.
DR   MGD; MGI:105380; Mcmd2.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding activity; IPI.
DR   GO; GO:0006268; P:DNA unwinding; IDA.
DR   GO; GO:0006334; P:nucleosome assembly; IDA.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger; ATP-binding; Cell cycle.
FT   ZN_FING     329    355       C4-TYPE (POTENTIAL).
FT   DOMAIN      473    680       MCM.
FT   NP_BIND     523    530       ATP (POTENTIAL).
FT   CONFLICT    506    506       N -> D (IN REF. 3).
FT   CONFLICT    521    521       L -> S (IN REF. 3).
FT   CONFLICT    648    648       D -> N (IN REF. 3).
SQ   SEQUENCE   904 AA;  102047 MW;  2C77E3BA1B0F2EEE CRC64;

  Query Match             85.1%;  Score 3941;  DB 1;  Length 904;
  Best Local Similarity   86.1%;  Pred. No. 6.2e-241;
  Matches  787;  Conservative   23;  Mismatches   62;  Indels   42;  Gaps    7;

Qy          2 ASSPA-QRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              |||||  ||||  :|||||||||||||  |||||||||||||||||||||||||:||||||
Db         10 ASSPARQRRRISDPLTSSPGRSSRRADALTSSPGRDLPPFEDESEGLLGTEGPMEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              |||||||||||| ||||| ||||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRPIPELDVYEAEGLALDDEDVEELTASQREAAERTMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTAMRRTRSALPASAASG------------AGTEDGEEDEQMIESIE 161
                 |||           |  |  |  |:                 ||||||||:||||||
Db        127 ---GLG-----------RMRRGLLYDSSEEDEERPARKRRHVERATEDGEEDEEMIESIE 172

Qy        162 NLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 221
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 NLEDLKGHSVRERVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 232

Qy        222 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 292

Qy        282 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNXCNFVLGPFCQSQNQEVKPGSC 341
              |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        293 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCSXCNFVLGPFCQSQNQEVKPGSC 352

Qy        342 PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEI 401
              ||||||||||:|||||||||||||||||||||| ||||||||||||||||  ||||
Db        353 PECQSAGPFEINMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEI 412

Qy        402 ELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQ 461
              ||||||||||||||||||||||:|||||||||||||||||||||||| ||||||
Db        413 ELTGIYHNNYDGSLNTANGFPVFATIILANHVAKKDNKVAVGELTDEDVKMITGLSKDQQ 472

Qy        462 IGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQF 521
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        473 IGEKIFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQF 532

Qy        522 LKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKM 581
              |||||||||||||||||||| :|||||||||||||||||||||||||||||||||||||||
Db        533 LKYIEKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKM 592

Qy        582 NDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEP 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 NDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEP 652

Qy        642 IISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGV 701
              ||||||:|||||||||||||||||||||||||||||::||| ||   ||||||||||
Db        653 IISRFDVLCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKKDEGLTNGGTLEPAMPNTYGV 712

Qy        702 EPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSH 761
              |||||||||||||||| |||||||||||:|||||||||||||||||||||||||||||||
Db        713 EPLPQEVLKKYIIYAKERVRPKLNQMDQDKVARMYSDLRKESMATGSIPITVRHIESMIR 772

Qy        762 GGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNE 818
                    ||    | |       :       :    : |:    ||||||||||||||||:
Db        773 MA--EAHARMHLRDYVMEDDVNMAIRVMMESFIDTQKFSVMRSMRKTFARYLSFRRDNND 830

Qy        819 LLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNK 878
              |||||||||||||||||||||||||||||:|||||:|||||||||||||||||:||: ||
Db        831 LLLFILKQLVAEQVTYQRNRFGAQQDTIEIPEKDLMDKARQINIHNLSAFYDSDLFKFNK 890

Qy        879 FSHDLKRKMILQQF 892
              || |||||:|||||
Db        891 FSRDLKRKLILQQF 904


RESULT 3
MCM2_XENLA
ID   MCM2_XENLA     STANDARD;      PRT;   886 AA.
AC   P55861;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA replication licensing factor MCM2 (X.MCM2).
GN   MCM2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
```

```
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocyte;
RX   MEDLINE=97074651; PubMed=8917078;
RA   Miyake S., Saito I., Kobayashi H., Yamashita S.;
RT   "Identification of two Xenopus laevis genes, xMCM2 and xCDC46, with
RT   sequence homology to MCM genes involved in DNA replication.";
RL   Gene 175:71-75(1996).
CC   -!- FUNCTION: ACTS AS A FACTOR THAT ALLOWS THE DNA TO UNDERGO A SINGLE
CC       ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR THE ENTRY IN
CC       S PHASE AND FOR CELL DIVISION.
CC   -!- SUBUNIT: ASSOCIATED AND INTERACTS WITH MCM5.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D63919; BAA09948.1; -.
DR   PIR; JC5085; JC5085.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger; ATP-binding; Cell cycle.
FT   ZN_FING     314    340       C4-TYPE (POTENTIAL).
FT   DOMAIN      458    665       MCM.
FT   NP_BIND     508    515       ATP (POTENTIAL).
FT   DOMAIN       49     52       POLY-GLU (ACIDIC).
FT   DOMAIN      125    128       POLY-GLU (ACIDIC).
SQ   SEQUENCE   886 AA;  100376 MW;  85B35A9EFA8174BD CRC64;

  Query Match           76.4%;  Score 3539.5;  DB 1;  Length 886;
  Best Local Similarity  78.3%;  Pred. No. 1.3e-215;
  Matches  704;  Conservative   57;  Mismatches   95;  Indels   43;  Gaps    8;

Qy         15 LTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRA 74
              : ::||   ||| |||||||||||||||||||: |     |||||||||||| ||||||
Db         10 IATSPRAGSRR-DALTSSPGRDLPPFEDESEGMFGDGVVPEEEEDGEELIGDAMERDYRP 68

Qy         75 IPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTGSWPGLGACAVGSCM 127
              | ||| || |||  |:||||:||||:||||       |   |   |
Db         69 ISELDRYEVEGLD-DEEDVEDLTASQREAAEQSMRMRDREMGRELG-------------- 113

Qy        128 TAMRR-----------TRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVS 176
                |||           | |     |  |     ||||:||||||||||:|||:||||||
Db        114 -RMRRGLLYDSDEEEEDRPARKRRMAERAAEGAPEEDEEMIESIENLEDMKGHTVREWVS 172

Qy        177 MAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYF 236
              ||  ||||:|||||||||||| |||||||:||||||||:||| |||||||||||||||
Db        173 MAATRLEIYHRFKNFLRTHVDEHGHNVFKEKISDMCKENKESLPVNYEDLAAREHVLAYF 232

Qy        237 LPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIR 296
              |||||||:|:|||||| |||| ||||||||   ||||||||||||||||||||||||||
Db        233 LPEAPAEMLKIFDEAAKEVVLVMYPKYDRIAREIHVRISHLPLVEELRSLRQLHLNQLIR 292

Qy        297 TSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEE 356
              |||||| |||||||||||||||||:|||| |||||||:||||||||| ||||:||||
Db        293 TSGVVTCCTGVLPQLSMVKYNCNKCNFILGPFFQSQNQEVRPGSCPECQSFGPFEINMEE 352

Qy        357 TIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLN 416
              |:||||||| |||||||||| ||||||||||||||||  ||||||||||||||||||||
Db        353 TVYQNYQRITIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTGIYHNNYDGSLN 412

Qy        417 TANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYG 476
              |||||||||||||||: |||:|||| |||||||| | :||||::|||:||||||||||
Db        413 TANGFPVFATVILANHITKKDDKVAVRELTDEDVKAIVALSKDERIGERIFASIAPSIYG 472

Qy        477 HEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTT 536
              ||||||| |||||||| ||||||||||||||||||||||||||||||:|||:|||:|||
Db        473 HEDIKRGLALALFGGEAKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYVEKVASRAVFTT 532

Qy        537 GQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQ 596
              ||||||| :|||||||::|||||||||| ||||||||||||||||||||||||||||||
Db        533 GQGASAVGLTAYVQRHPVTKEWTLEAGALVFADRGVCLIDEFDKMNDQDRTSIHEAMEQQ 592

Qy        597 SISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTV 656
              ||||||||||||||||||||:||||||||||||||||||||||:||||||||||||||
Db        593 SISISKAGIVTSLQARCTVIAASNPIGGRYDPSLTFSENVDLTEPIVSRFDILCVVRDTV 652

Qy        657 DPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYA 716
              |||||||||||||| ||::||||:|:   :||| ||| |:|||:||| ||||||||||:||
Db        653 DPVQDEMLARFVVSSHIKHHPSSKD---IANGDAAEFALPNTFGVEALPQEVLKKYIMYA 709

Qy        717 KERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDR 776
              ||:: |||||||||||||:|||||||||||||||||||||||||       ||    | |
Db        710 KEKIRPKLNQMDQDKVAKIYSDLRKESMATGSIPITVRHIESMIRMA--EAHARMHLRDY 767

Qy        777 RRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLVAEQVT 833
                        :       |     : |:    ||||||||||:||||||||||:||||:|||||
Db        768 VVEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLAFRRDNNELLLFVLKQLIAEQVT 827

Qy        834 YQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892
              |||||:|||||||||||||||||||||||||||||||:||:||||:||:|:|:|:|||
Db        828 YQRNRYGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSDLFKMNKFTHDVKKKLIIQQF 886


RESULT 4
MCM2_DROME
ID   MCM2_DROME     STANDARD;      PRT;   887 AA.
AC   P49735; Q9VHU2;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA replication licensing factor MCM2.
GN   MCM2 OR CG7538.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye-antennal disk;
RX   MEDLINE=95347580; PubMed=7622035;
RA   Treisman J.E., Follette P.J., O'Farrell P.H., Rubin G.M.;
RT   "Cell proliferation and DNA replication defects in a Drosophila MCM2
RT   mutant.";
RL   Genes Dev. 9:1709-1715(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
```

```
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: ACTS AS A FACTOR THAT ALLOWS THE DNA TO UNDERGO A
CC       SINGLE ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR DNA
CC       REPLICATION AND CELL PROLIFERATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L42762; AAB36617.1; -.
DR   EMBL; AE003678; AAF54207.1; -.
DR   FlyBase; FBgn0014861; Mcm2.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger; ATP-binding; Cell cycle.
FT   ZN_FING      314    340       C4-TYPE (POTENTIAL).
FT   DOMAIN       458    665       MCM.
FT   NP_BIND      508    515       ATP (POTENTIAL).
SQ   SEQUENCE   887 AA;  100414 MW;  26A7092109F09CBE CRC64;

  Query Match             57.2%;  Score 2649;  DB 1;  Length 887;
  Best Local Similarity   57.7%;  Pred. No. 2e-159;
  Matches  527;  Conservative  134;  Mismatches  179;  Indels   74;  Gaps   12;

Qy         12 NDPLTSSPGRSSRRTD--ALTSSPGRDLPPFEDESEGLLGTEGPLE--EEEDGEELIGDG 67
              : |  ::|  :: | |  |  :||  |  |||:| | :|| :   :  |||||||| ||
Db          5 SSPPPNTPSDAAERRDLRAAMTSPVGDFEPFENEDE-ILGDQTVRDEAEEEDGEELFGDN 63

Qy         68 MERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCM 127
              || ||| :|||| |:   |  |::|  |::   | ||:
Db         64 MENDYRPMPELDHYD-PALLDDEDDFSEMSQGDRFAAE---------------------- 100

Qy        128 TAMRRTRSALPASAAS---GAGTEDGE--------------------EDEQMIESIENLE 164
              : |||   |           | |  | |                   || :|:|||||||
Db        101 SEMRRRDRAAGIHRDDRDLGFGQSDDEDDVGPRAKRRAGEKAAVGEVEDTEMVESIENLE 160

Qy        165 DLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYE 224
              | |||| :|||| ||| || :||::|||| ||  |   :::||  ||::|  | ||:|
Db        161 DTKGHSTKEWVSMLGPRTEIANRFQSFLRTFVDERGAYTYRDRIRRMCEQNMSSFVVSYT 220

Qy        225 DLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELR 284
              ||| :|||||||||||| ::|:|||: | ::||:::| |:|:|  |||||| |||:||||
Db        221 DLANKEHVLAYFLPEAPFQMLEIFDKVAKDMVLSIFPTYERVTTEIHVRISELPLIEELR 280

Qy        285 SLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPEC 344
              : |:||||||:|| |||||: |||||||||::||:| || :||||| |||| |:||||||||
Db        281 TFRKLHLNQLVRTLGVVTATTGVLPQLSVIKYDCVKCGYVLGPFVQSQNTEIKPGSCPEC 340

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
              || ||| :|||:|:|:|||:| :||||||:: | |:||||| |||||| | |  |||:|:|
Db        341 QSTGPFSINMEQTLYRNYQKITLQESPGRIPAGRIPRSKDVILLADLCDQCKPGDELEVT 400

Qy        405 GIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGE 464
              ||| |||||||||  |||||||||:|||  ||:|  |  |||||:  |  |||| :| |
Db        401 GIYTNNYDGSLNTDQGFPVFATVIIANHVVVKDSKQVVQSLTDEDIATIQKLSKDPRIVE 460

Qy        465 KIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKY 524
              :: ||:|||||||: |||  |||||||| |||| |||||||||:|:|||||||||||||
Db        461 RVVASMAPSIYGHDYIKRALALALFGGESKNPGEKHKVRGDINLLICGDPGTAKSQFLKY 520

Qy        525 IEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQ 584
               |||: ||:|||||||||| :||||:|:|||||||||||||||||:|||||||||||||
Db        521 TEKVAPRAVFTTGQGASAVGLTAYVRRNPVSREWTLEAGALVLADQGVCLIDEFDKMNDQ 580

Qy        585 DRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIIS 644
              |||||||||||||||||||||||||||||||||||||||||||||||:||||||:|:|||:|
Db        581 DRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSMTFSENVNLSEPILS 640

Qy        645 RFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPL 704
              |||:||||:|  ||:||: ||:||| ||::|||| :|:  |           |    |: :
Db        641 RFDVLCVVKDEFDPMQDQQLAKFVVHSHMKHHPSEEEQPELEE--------PQLKTVDEI 692

Qy        705 PQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGG 764
              ||::|::||:|||| : |||  :|:||:||||: ||:|| ||||:||||||||:
Db        693 PQDLLRQYIVYAKENIRPKLTNIDEDKIAKMYAQLRQESFATGSLPITVRHIESVIRMS- 751

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAG-----ELHRHTEVQRH---RSMRKTFARYLSFRRDN 816
              ||    |     |::    |        |    | |:    : || || :||||::|:
Db        752 -EAHARMHL-----RENVMEADVSMAIRMMLESFIEAQKFSVMKKMRSTFQKYLSFQKDH 805

Qy        817 NELLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRM 876
              :||| |||:||  :|: | | : |     :|: |:||:::|:|::| ||  ||:|:|||
Db        806 SELLFFILRQLTLDQLAYIRCKDGPGATHVEIMERDLIERAKQLDIVNLKPFYESDLFRT 865

Qy        877 NKFSHDLKRKMILQ 890
              | ||:| ||::|||
Db        866 NGFSYDPKRRIILQ 879


RESULT 5
CC19_SCHPO
ID   CC19_SCHPO     STANDARD;      PRT;   830 AA.
AC   P40377;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA replication licensing factor cdc19 (Cell division control protein
DE   19).
GN   CDC19 OR NDA1 OR SPBC4.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95181580; PubMed=7876346;
RA   Forsburg S.L., Nurse P.;
RT   "The fission yeast cdc19+ gene encodes a member of the MCM family of
RT   replication proteins.";
RL   J. Cell Sci. 107:2779-2788(1994).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94129084; PubMed=8298187;
RA   Miyake S., Okishio N., Samejima I., Hiraoka Y., Toda T.,
RA   Saitoh I., Yanagida M.;
RT   "Fission yeast genes nda1+ and nda4+, mutations of which lead to
RT   S-phase block, chromatin alteration and Ca2+ suppression, are members
RT   of the CDC46/MCM2 family.";
RL   Mol. Biol. Cell 4:1003-1015(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN DNA REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08048; AAC48930.1; -.
DR   EMBL; S68468; AAC60569.1; -.
DR   EMBL; AL121863; CAB58403.1; -.
DR   PIR; B48723; B48723.
DR   GeneDB_SPombe; SPBC4.04c; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger; ATP-binding; Cell cycle.
FT   ZN_FING      334    360       C4-TYPE (POTENTIAL).
FT   DOMAIN       484    691       MCM.
FT   NP_BIND      534    541       ATP (POTENTIAL).
SQ   SEQUENCE   830 AA;  92831 MW;  A5BF0AA29661BD66 CRC64;

  Query Match             42.1%;  Score 1951.5;  DB 1;  Length 830;
  Best Local Similarity   48.2%;  Pred. No. 1.9e-115;
  Matches  418;  Conservative  116;  Mismatches  238;  Indels   95;  Gaps   12;

Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLE----- 55
              | |   : || :: |      ||     |: |    | | ||:   |  |  :|
Db          1 MDSFRKRGRRDSESLPFESENSSLGATPLSLPPSSPPPEFSDEAAEAL-VEEDIEDLDGE 59

Qy         56 ----EEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTV 111
                   |:|:||:| |:|||||:   ||| |: | |  || |:|||    | | |
Db         60 ALDVEDEEGEDLFGEGMERDYQQNLELDRYDIEELD-DDNDLEELDIGARRAVD------ 112

Qy        112 TGSWPGLGACAVGSCMTAMRRTRSALPASAA----------------SGAGT-------- 147
                              :||    | |:|                  |  ||
Db        113 ----------------ARLRRRDIELDAAAGRTKPAAFLQDEDDDLDSNLGTGFTRHRHR 156

Qy        148 -------EDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHG 200
                    |  ||      :|:: |:|  |: |||::   |  |   ||||  : | :|
Db        157 IYDEYSPNVGALDESGELPLESIADVKADSIAEWVTLDPVRRTIAREFKNFLLEYTDENG 216

Qy        201 HNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMY 260
               :|:  ||  : : | |||:|||  |   : :|||||  ||| : :|||  |||  |  |
Db        217 TSVYGNRIRTLGEVNAESLMVNYAHLGESKPILAYFLANAPAPIFRIFDRVALEATLLHY 276

Qy        261 PKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNK 320
               | |:|| : |||||::||    || ||| ||| |:| |||||  ||: |||  ::: | |
Db        277 PDYERIHSDIHVRITNLPTCFTLRDLRQSHLNCLVRVSGVVTRRTGLFPQLKYIRFTCTK 336

Qy        321 CNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLP 380
               |  |||| |  : |||   |  | | ||| :| | |:| ||||| :||||| | : |||
Db        337 CGATLGPFFQDSSVEVKISFCHNCSSRGPFVINSERTVYNNYQRITLQESPGTVPSGRLP 396

Qy        381 RSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAK----- 435
               | :: |||||||    |:|||:|||| ||:| |||| |||||||:| |||:::
Db        397 RHREVILLADLVDVAKPGEEIDVTGIYRNNFDASLNTKNGFPVFATIIEANHISQLDGSG 456

Qy        436 -KDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPK 494
                |:  ::  |||:: : | :|:|   |  :| ||:|||||||  ||   | ||||| ||
Db        457 NTDDDFSLSRLTDDEEREIRALAKSPDIHNRIIASMAPSIYGHRSIKTAIAAALFGGVPK 516

Qy        495 NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPV 554
               |  ||||:|||||||| ||||||||||||||:|| : ||:| |||||||| :|| |:: |:
Db        517 NINGKHKIRGDINVLLLGDPGTAKSQFLKYVEKTAHRAVFATGQGASAVGLTASVRKDPI 576

Qy        555 SREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCT 614
               : ||||| |||||||:|||||||||||||||||||||||||||||||||||:||||||
Db        577 TNEWTLEGGALVLADKGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTTLQARCT 636

Qy        615 VIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVR 674
               :|||||||||||: :: |::||:|||||:|||||| ||:|||:|  || || ||| ||:|
Db        637 IIAAANPIGGRYNTTIPFNQNVELTEPILSRFDILQVVKDTVNPEIDEQLANFVVSSHIR 696

Qy        675 HHPSNKEEEGLANGSAAEPAM------PNTYGVE--PLPQEVLKKYIIYAKERVHPKLNQ 726
                ||            | :| |      |   |::  |:||::|:||| :|:|:| |:| |
Db        697 SHP------------AFDPNMDVLKKVPTETGIDAKPIPQDLLRKYIHFAREKVFPRLQQ 744

Qy        727 MDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRD 786
               ||::|::::|||:|:||:|||| ||||||:||      |    | :  |  |
Db        745 MDEEKISRLYSDMRRESLATGSYPITVRHLESAIRLS--EAFAKMQLSEFVRPSHIDKAI 802

Qy        787 AGELHRHTEVQR---HRSMRKTFARYL 810
                 :      |:    ||: :|||:||
Db        803 QVIIDSFVNAQKMSVKRSLSRTFAKYL 829


RESULT 6
MCM2_YEAST
ID   MCM2_YEAST     STANDARD;      PRT;   868 AA.
AC   P29469;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA replication licensing factor MCM2 (Minichromosome maintenance
DE   protein 2).
GN   MCM2 OR YBL023C OR YBL0438.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
```

```
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91257581; PubMed=2044961;
RA   Yan H., Gibson S., Tye B.K.;
RT   "Mcm2 and Mcm3, two proteins important for ARS activity, are related
RT   in structure and function.";
RL   Genes Dev. 5:944-957(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95242843; PubMed=7725803;
RA   van Dyck L., Jonniaux J.-L., Barreiros T.D.M., Kleine K., Goffeau A.;
RT   "Analysis of a 17.4 kb DNA segment of yeast chromosome II
RT   encompassing the ribosomal protein L19 as well as proteins with
RT   homologies to components of the hnRNP and snRNP complexes and to the
RT   human proliferation-associated p120 antigen.";
RL   Yeast 10:1663-1673(1994).
RN   [3]
RP   SEQUENCE OF 1-85 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94103216; PubMed=8276800;
RA   van Dyck L., Pearce D.A., Sherman F.;
RT   "PIM1 encodes a mitochondrial ATP-dependent protease that is required
RT   for mitochondrial function in the yeast Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 269:238-242(1994).
CC   -!- FUNCTION: NECESSARY FOR CELL GROWTH. POSSIBLY PLAYS A DIRECT
CC       ROLE IN THE INITIATION OF CHROMOSOMAL DNA REPLICATION BY
CC       INTERACTING DIRECTLY WITH AUTONOMOUSLY REPLICATING SEQUENCES
CC       (ARS). ITS ARS SPECIFICITY IS NOT IDENTICAL TO THE HIGHLY SIMILAR
CC       MCM3 PROTEIN.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X77291; CAA54503.1; -.
DR   EMBL; X53539; CAA37615.1; -.
DR   EMBL; Z35784; CAA84842.1; -.
DR   EMBL; X74544; CAA52635.1; -.
DR   PIR; S45757; S45757.
DR   GK; P29469; -.
DR   SGD; S0000119; MCM2.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005656; C:pre-replicative complex; IDA.
DR   GO; GO:0006267; P:pre-replicative complex formation and maint. . .; IPI.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Zinc-finger; ATP-binding; Cell cycle.
FT   ZN_FING      341    367       C4-TYPE (POTENTIAL).
FT   DOMAIN       493    700       MCM.
FT   NP_BIND      543    550       ATP (POTENTIAL).
FT   VARIANT      392    392       E -> K (IN STRAIN MCM2-1).
FT   MUTAGEN      364    364       C->Y,F,S,H: LOSS OF ACTIVITY.
FT   MUTAGEN      367    367       C->Y,F,S,H: LOSS OF ACTIVITY.
FT   CONFLICT     164    164       S -> T (IN REF. 1).
FT   CONFLICT     172    173       MD -> IH (IN REF. 1).
FT   CONFLICT     529    529       G -> P (IN REF. 1).
FT   CONFLICT     578    578       A -> R (IN REF. 1).
FT   CONFLICT     583    583       D -> H (IN REF. 1).
FT   CONFLICT     712    712       E -> Q (IN REF. 1).
FT   CONFLICT     733    747       MISSING (IN REF. 1).
FT   CONFLICT     859    868       RSFAIYTLGH -> SLSQFIPWVTKTLLFLRISGYEDKKFS
FT                                 VSIHVLAILFSIYKFPLFFV (IN REF. 1).
SQ   SEQUENCE   868 AA;  98779 MW;  F38FF682581B0EC0 CRC64;

  Query Match             39.1%;  Score 1812;  DB 1;  Length 868;
  Best Local Similarity   46.3%;  Pred. No. 1.3e-106;
  Matches  384;  Conservative  112;  Mismatches  222;  Indels  112;  Gaps    9;

Qy          3 SSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEE 62
              ||| |  ||     |||            ||    |  :|     :     :||: :  :
Db         23 SSPQQHFRGGMNPVSSP----------IGSPDMINPEGDDNEVDDVPDIDEVEEQMNEVD 72

Qy         63 LIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACA 122
              |: | |  || |    | |: :   :|| : :||: | |   |
Db         73 LMDDNMYEDYAADHNRDRYDPD--QVDDREQQELSLSERRRID----------------- 113

Qy        123 VGSCMTAMRRTRSALPASAASGAGTEDGEEDEQ--------------------MIESIEN 162
                  |   |  |  : |      | |::||                       | :||
Db        114 ------AQLNERDRLLRNVA----YIDDEDEEQEGAAQLDEMGLPVQRRRRRRQYEDLEN 163

Qy        163 LED----------------------LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHG 200
               :|                     :| :|  ||::       |    |:||  : |  |
Db        164 SDDDLLSDMDIDPLREELTLESLSNVKANSYSEWITQPNVSRTIARELKSFLLEYTDETG 223

Qy        201 HNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMY 260
              :|:  ||  : : | ||| |||  ||  : :|| || : | |:|:|||  |:|     |
Db        224 RSVYGARIRTLGEMNSESLEVNYRHLAESKAILALFLAKCPEEMLKIFDLVAMEATELHY 283

Qy        261 PKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNK 320
              | | || : ||||||  | :  || ||: :|: |:| :|||  ||| |||  ||:|| |
Db        284 PDYARIHSEIHVRISDFPTIYSLRELRESNLSSLVRVTGVVTRRTGVFPQLKYVKFNCLK 343

Qy        321 CNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLP 380
               |  :|||| |  |:|::   |  |:| ||| || |:|:|:|||: :||:|| |   |||
Db        344 CGSILGPFFQDSNEEIRISFCTNCKSKGPFRVNGEKTVYRNYQRVTLQEAPGTVPPGRLP 403

Qy        381 RSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKV 440
              | :: |||||||||    |:|:|:|||| ||||||:|| |||||||:| || : :::
Db        404 RHREVILLADLVDVSKPGEEVEVTGIYKNNYDGNLNAKNGFPVFATIIEANSIKRREGNT 463

Qy        441 A--------VGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
              |        |   |:|: :    :|:|: | :|| :|:|||||| |||   | :||||
Db        464 ANEGEEEGLDVFSWTEEEEREFRKISRDRGIIDKIISSMAPSIYGHRDIKTAVACSLFGGV 523

Qy        493 PKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRH 552
              |||  ||| :|||||||| |||||||| |||:|| : ||:| |||||||  :|| |::
Db        524 PKNVNGKHSIRGDINVLLLGDPGTAKSQILKYVEKTAHRAVFATGQGASAVGLTASVRKD 583

Qy        553 PVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQAR 612
              |:::|||| |||||||:|||||||||||||||||||||||||||||||||:|||
Db        584 PITKEWTLEGGALVLADKGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTTLQAR 643

Qy        613 CTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSH 672
              |::|||||| |||| :|  ::|| |||||:|||||||||| ||   || || ||| ||
Db        644 CSIIAAANPNGGRYNSTLPLAQNVSLTEPILSRFDILCVVRDLVDEEADERLATFVVDSH 703

Qy        673 VRHHPSNKE----EEGLANGSAA------------------EPAMPNTYGVEPLPQEVL 709
              || || | |     ||   || :|                  :       : |:|||:|
Db        704 VRSHPENDEDREGEELKNNGESAIEQGEDEINEQLNARQRRLQRQRKKEEEISPIPQELL 763

Qy        710 KKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESM 759
               ||| ||: :::|||:||| |||:::|:|||:||::||| ||||||:||:
Db        764 MKYIHYARTKIYPKLHQMDMDKVSRVYADLRRESISTGSFPITVRHLESI 813


RESULT 7
MCM4_HUMAN
ID   MCM4_HUMAN     STANDARD;      PRT;   863 AA.
AC   P33991; Q99658;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE   DNA replication licensing factor MCM4 (CDC21 homolog) (P1-CDC21).
GN   MCM4 OR CDC21.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hu B.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-712 FROM N.A.
RX   MEDLINE=98126438; PubMed=9465298;
RA   Connelly M.A., Zhang H., Kieleczawa J., Anderson C.W.;
RT   "The promoters for human DNA-PKcs (PRKDC) and MCM4: divergently
RT   transcribed genes located at chromosome 8 band q11.";
RL   Genomics 47:71-83(1998).
RN   [3]
RP   SEQUENCE OF 1-23 FROM N.A.
RX   MEDLINE=97430835; PubMed=9284934;
RA   Ladenburger E.M., Fackelmayer F.O., Hameister H., Knippers R.;
RT   "MCM4 and PRKDC, human genes encoding proteins MCM4 and DNA-PKcs, are
RT   close neighbours located on chromosome 8q12-->q13.";
RL   Cytogenet. Cell Genet. 77:268-270(1997).
RN   [4]
RP   SEQUENCE OF 440-863 FROM N.A.
RC   TISSUE=Cervix;
RX   MEDLINE=94089373; PubMed=8265339;
RA   Hu B., Burkhart R., Schulte D., Musahl C., Knippers R.;
RT   "The P1 family: a new class of nuclear mammalian proteins related to
RT   the yeast Mcm replication proteins.";
RL   Nucleic Acids Res. 21:5289-5293(1993).
CC   -!- FUNCTION: INVOLVED IN THE CONTROL OF DNA REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74794; CAA52801.1; -.
DR   EMBL; U63630; AAC52018.1; -.
DR   EMBL; U90415; AAB51723.3; -.
DR   PIR; S65954; S65954.
DR   Genew; HGNC:6947; MCM4.
DR   GK; P33991; -.
DR   MIM; 602638; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0005524; F:ATP binding activity; NAS.
DR   GO; GO:0006260; P:DNA replication; NAS.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      458    667       MCM.
FT   NP_BIND     510    517       ATP (POTENTIAL).
FT   CONFLICT     62     62       T -> P (IN REF. 2).
FT   CONFLICT    206    206       Q -> P (IN REF. 2).
FT   CONFLICT    650    650       M -> L (IN REF. 2).
SQ   SEQUENCE   863 AA;  96610 MW;  50010575C41BA90C CRC64;

  Query Match            18.9%;  Score 877;  DB 1;  Length 863;
  Best Local Similarity  30.4%;  Pred. No. 1.3e-47;

    Matches  246;  Conservative  129;  Mismatches  280;  Indels  154;  Gaps   23;

Qy          3 SSPAQRRRGND--------PLTSSPG---RSSRRTDALTSSPGRDLPPFEDESEGLLGTE 51
              |||:|||||  |         |: :|||   :|:   | | |||
Db         31 SSPSQRRRGEDSTSTGELQPMPTSPGVDLQSTAAQDVLFSSP----------------- 72

Qy         52 GPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT----ASRREAA--D 105
                                    |::    :  : || :    ||    :|| |
Db         73 ----------------------PQM---HSSAIPLDFDVSSPLTYGTPSSRVEGTPRS 105

Qy        106 GPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
              |  ||     | ||:   |  :               : ||  ||    || :
Db        106 GVRGTPVRQRPDLGSAQKGLQVDLQ----------SDGAAAEDIVASEQSL-------- 146

Qy        166 LKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVD---SHGHNV--------FKERISDM 211
                |  :  |   |::|  :       |: ||:  :|       ||        : :|: ::
Db        147 --GQKLVIWGTDVNVAACK----ENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEI 200

Qy        212 CKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNH-I 270
                : | || | : : :  |   |   | |::  || |  |:    ||  | |  | |
Db        201 NVIGEQFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYP--DSILEHQI 258

Qy        271 HVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 330
              ||  :   : :|:|    ::|||  ||:|   : ::|::    : |  |
Db        259 QVRPFNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTT-RVEM 317

Qy        331 SQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLA- 389
              : :  :|  |  | :      :  ::: : | |::||||  : | : | :   || |
Db        318 DRGRIAEPSVCGRCHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHT--VILFAH 375

Qy        390 -DLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATVILANHVAKKDNKVAV 442
               ||||    || : :||||       |        :|   |: | |   |  | | | |
Db        376 NDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVSNVKSVYKTHIDVIHYRKTDAKRLH 430

Qy        443 GELTDED----------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
              |   ||:          |:::  ||:   | |:: :::||||| |||||:|  | ||||
Db        431 G--LDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGT 488

Qy        493 PK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQ 550
              |  :  |: | | :||:|||||||||:||| |:|:  :  |  :|:|:|:|:||| :||||
Db        489 RKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVM 548

Qy        551 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 610
              : | :|:  |: ||||||:| |:| |||||||||:  |: :|| ||||::||:||||:  |
Db        549 KDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLN 608

Qy        611 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVG 670
              || :|:|||||||  :::|  |  ||: |   ::||||:: :: |  |   |  ||  :|
Db        609 ARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVA 668

Qy        671 SHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD 730
              : :    :|                | |   ||| || ||   : |:|::
Db        669 LYYQSEEQAEE-------------------ELLDMAVLKDYIAYAHSTIMPRLSEEASQ 708

Qy        731 KVAKMYSDLRKESMATGSIPITVRHIESM 759
              : : | |:||   : | :    | :||:
Db        709 ALIEAYVDMRKIGSSRGMVSAYPRQLESL 737


RESULT 8
MCM4_XENLA
ID   MCM4_XENLA     STANDARD;      PRT;   863 AA.
AC   P30664; Q91679;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA replication licensing factor MCM4 (CDC21 homolog) (P1-CDC21)
DE   (X.MCM4).
GN   MCM4.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
```

```
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96183193; PubMed=8605878;
RA   Coue M., Kearsey S.E., Mechali M.;
RT   "Chromotin binding, nuclear localization and phosphorylation of
RT   Xenopus cdc21 are cell-cycle dependent and associated with the
RT   control of initiation of DNA replication.";
RL   EMBO J. 15:1085-1097(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97057222; PubMed=8901561;
RA   Hendrickson M., Madine M., Dalton S., Gautier J.;
RT   "Phosphorylation of MCM4 by cdc2 protein kinase inhibits the activity
RT   of the minichromosome maintenance complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:12223-12228(1996).
RN   [3]
RP   SEQUENCE OF 513-588 FROM N.A.
RX   MEDLINE=93087163; PubMed=1454522;
RA   Coxon A., Maundrell K., Kearsey S.E.;
RT   "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT   early step of chromosome replication.";
RL   Nucleic Acids Res. 20:5571-5577(1992).
CC   -!- FUNCTION: INVOLVED IN THE CONTROL OF DNA REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29178; AAB01680.1; -.
DR   EMBL; U46131; AAA91232.1; -.
DR   EMBL; Z15033; CAA78751.1; -.
DR   PIR; S64720; S64720.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      458    667       MCM.
FT   NP_BIND     510    517       ATP (POTENTIAL).
FT   CONFLICT     73     73       V -> A (IN REF. 2).
FT   CONFLICT    740    741       RA -> LS (IN REF. 2).
SQ   SEQUENCE   863 AA;  97178 MW;  9A47003097EEDE2E CRC64;

  Query Match             18.9%;  Score 874.5;  DB 1;  Length 863;
  Best Local Similarity   28.1%;  Pred. No. 1.9e-47;
  Matches  264;  Conservative  145;  Mismatches  358;  Indels  173;  Gaps   25;

Qy          3 SSPAQRR----RGNDPLT------SSPGRSSRRTDALTS---------SPGRDLPPFEDE 43
              |:|::||    ||::| |       ||    |||:  ||         ||  |:
Db          6 STPSRRRNKRGRGSNPPTPHGEEVQSPPSQRRRTEDSTSIGELLPMPTSPSGDV------ 59

Qy         44 SEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREA 103
                     :   |:||:          :|   :     | |         :|| |
Db         60 ------------QSPSGQELLFSS------PVPSRHSAHQSELDLSSPLTYGTPSSRVEG 101

Qy        104 A--DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIE 161
                 |  ||     | ||:          |: :               |   |:   | :
Db        102 TPRSGIRGTPARQRPDLGSA---------RKVKQV------------DLHSDQPAAEELV 140

Qy        162 NLEDLKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVDSHGHN-----------VFKER 207
                  |   |  :  |   |::|        :|: |::  :|                :: :|

Db        141 TSEQSLGQKLVIWGTDVNVA----TCKEKFQRFVQRFIDPSAKEEDNVGLDLNEPIYMQR 196

Qy        208 ISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRIT 267
              : ::       | :: : |    :  |   |   | |::  || || |:    ||  | |
Db        197 LEEINVVGDPFLNIDCDHLRNFDQDLYRQLVCYPQEVIPTFDMAANEIFFERYP--DSIL 254

Qy        268 NH-IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLG 326
               | | ||  :      :|||    ::|||  ||:|   : ::|::    : |  | |
Db        255 EHQIQVRPYNALKTRNMRSLNPEDIDQLITISGMVIRTSQIIPEMQEAFFKCQVCAFTTR 314

Qy        327 PFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAI 386
                  : :  :|  |  | :     :    ::: : | |::|||  : | : | :
Db        315 VEI-DRGRIAEPSVCKHCNTTHSMALIHNRSMFSDKQMIKLQESPEDMPAGQTPHTTILY 373

Qy        387 LLADLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATVILANHVAKKDNKV 440
               ||||    || : :||||       |          |   |: | |   |  | |:|
Db        374 GHNDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVRNVKSVYKTHIDVIHYRKTDSKR 428

Qy        441 AVG--------ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
               |          |:| | |:  |:    | |:: |::||||| |||||:|  | ||||
Db        429 LHGIDEDTEQKLFTEERVAMLKELAAKPDIYERLAAALAPSIYEHEDIKKGILLQLFGGT 488

Qy        493 PK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQ 550
              |  :  |: | | ::|:|||||||:|:|| |:|:  :  |  :|:|:|:|:|| :||||
Db        489 RKDFSHTGRGKFRAEVNILLCGDPGTSKSQLLQYVFNLVPRGQYTSGKGSSAVGLTAYVM 548

Qy        551 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 610
              : | :|:  |: |||||:| |:| ||||||||:  |: :|| ||||::||:||||:  |
Db        549 KDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLN 608

Qy        611 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVG 670
               || :|:|||||:  :::|  |  ||: |   ::||||:: :: |  |   |  ||  :|
Db        609 ARTSVLAAANPVESQWNPKKTTIENIQLPHTLLSRFDLIFLMLDPQDEAYDRRLAHHLVA 668

Qy        671 SHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD 730
              : :     |||                      |   ||| || ||:  |:|:|::
Db        669 LYYQSEEQMKEEH--------------------LDMAVLKDYIAYARTYVNPRLSEEASQ 708

Qy        731 KVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPR----- 785
              : : |  :||      | :    | :||:                  ||   |   |
Db        709 ALIEAYVSMRKIGSGRGMVSAYPRQLESLI--------------RRAEAHAKVRFSNKV 753

Qy        786 ------DAGELHRHTEVQRHRSMRK--------TFARYLSFRRDNNELLLFILKQLVAEQ 831
                   :|  |||    |    |          |     : |:   ||   :||:|:  :
Db        754 ETIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATARKRKEELAQ-VLKKLIQSK 812

Qy        832 -----VTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLS 866
                  : ||:     :  :     ||: |:|    :| |:
Db        813 GKTPALKYQQLFEDLRGQSDAAITKDMFDEA----LHALA 848


RESULT 9
MCM3_SCHPO
ID   MCM3_SCHPO     STANDARD;      PRT;   879 AA.
AC   P30666;
DT   01-APR-1993 (Rel. 25, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA replication licensing factor mcm3 (Minichromosome maintenance
DE   protein 3).
GN   MCM3 OR SPCC1682.02C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP011;
RX   MEDLINE=98371232; PubMed=9705504;
RA   Sherman D.A., Forsburg S.L.;
RT   "Schizosaccharomyces pombe Mcm3p, an essential nuclear protein,
RT   associates tightly with Nda4p (Mcm5p).";
```

```
RL   Nucleic Acids Res. 26:3955-3960(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   SEQUENCE OF 359-434 FROM N.A.
RX   MEDLINE=93087163; PubMed=1454522;
RA   Coxon A., Maundrell K., Kearsey S.E.;
RT   "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT   early step of chromosome replication.";
RL   Nucleic Acids Res. 20:5571-5577(1992).
CC   -!- FUNCTION: PROBABLY INVOLVED IN THE INITIATION OF DNA REPLICATION.
CC   -!- SUBUNIT: ASSOCIATES STRONGLY WITH NDA4.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF063864; AAC32263.1; -.
DR   EMBL; AL031525; CAA20668.1; -.
DR   EMBL; Z15034; CAA78752.1; -.
DR   PIR; S26642; S26642.
DR   PIR; T41059; T41059.
DR   GeneDB_SPombe; SPCC1682.02c; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 2.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      306    513       MCM.
FT   NP_BIND     356    363       ATP (POTENTIAL).
SQ   SEQUENCE   879 AA;  97481 MW;  876051CC7DE2504A CRC64;

  Query Match             18.9%;  Score 873;  DB 1;  Length 879;
  Best Local Similarity   34.1%;  Pred. No. 2.4e-47;

    Matches  219;  Conservative  108;  Mismatches  225;  Indels   90;  Gaps   13;

Qy        188 FKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQI 247
              |: :|    |     :::| |  |     :  |:|| ::|      ||  :   | | ::
Db         17 FQEYLEHDTDDANVTLYQEAILRMLNMGQRRLIVNIDELRDYNRELADGVLLKPLEYVEP 76

Qy        248 FDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEEL----------------RSLRQLHL 291
              ||||   ||  :               : |  | ::|                |:|| :||
Db         77 FDEALRNVVSTLIDP----------VVHKDLKDKLFYVGFRGSFGDHHVNPRTLRAMHL 125

Qy        292 NQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC-------PEC 344
              |::|   |:|| |: | |::    : |         | |  :  :  |          :
Db        126 NKMISLEGIVTRCSFVRPKVIKSVHYCEATK--RHHFKQYADATMNGGLSFQSTVYPTQD 183

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
              ::  |  :   : ::::| | :|| | :   :|||| | :|  ||||:   || : :
Db        184 ENGNPLSIEFGFSTFRDHQSISLQEMPERAPPGQLPRSIDILLDDDLVDTVKPGDRVNIV 243

Qy        405 GIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVG-------ELTDEDVKMITSLS 457
              | | :   ||  : |    | ||:|||:|    ||  :|       ::|| |:: |  |:
Db        244 GQYRSM--GSKTSGNTSATFRTVLLANNVVLLGNKPGLGNVGGGALDITDADIRNINKLA 301

Qy        458 KDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKN-PGGKHKVRGDINVLLCGDPGT 516
              : : : | :  |:||||||:| :|:   | | ||  ||   | | :|||||:|: ||| |
Db        302 RKKNVFELLSTSLAPSIYGYEYVKQAILLLLLGGTEKNLTNGTH-IRGDINILMVGDPST 360

Qy        517 AKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLID 576
              |||| |:::   :  || |||:|:| | :|| |     : |  |||||:|||||||  ||
Db        361 AKSQLLRFVLNTAPLAIATTGRGSSGVGLTAAVTTDKETGERRLEAGAMVLADRGVVCID 420

Qy        577 EFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENV 636
              |||||:| || :||| ||||:::|:|||| ||| ||||:|||||||| |:||      :|:
Db        421 EFDKMSDIDRVAIHEVMEQQTVTIAKAGIHTSLNARCSVIAAANPIYGQYDIRKDPHQNI 480

Qy        637 DLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEE--------EGLANG 688
              | : ::||||:| :| | :|  :|  |:  |:  | |: |   |           : |
Db        481 ALPDSMLSRFDLLFIVTDDIDDKKDRALSEHVLRMH-RYLPPGVEPGTPVRDSLNSVLNV 539

Qy        689 SAAEPAMPNTYGVEPLPQEV--------------------------LKKYIIYAKERV 720
              |   |  :|  ||   |||                        ::||| ||| |:
Db        540 GATNAAGVSTENVE---QEVETPVWETFSSLLHANARTKKKELLNINFVRKYIQYAKSRI 596

Qy        721 HPKLNQMDQDKVAKMYSDLRKESMATG---SIPITVRHIESM 759
              || |||   :  :|  || : :      : |:| | :|::
Db        597 HPILNQATAEYITNIYCGLRNDDLQGNQRRTSPLTARTLETL 638


RESULT 10
MCM4_MOUSE
ID   MCM4_MOUSE     STANDARD;      PRT;   862 AA.
AC   P49717; O89056;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA replication licensing factor MCM4 (CDC21 homolog) (P1-CDC21).
GN   MCM4 OR MCMD4 OR CDC21.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95334361; PubMed=7610039;
RA   Kimura H., Takizawa N., Nozaki N., Sugimoto K.;
RT   "Molecular cloning of cDNA encoding mouse Cdc21 and CDC46 homologs
RT   and characterization of the products: physical interaction between
RT   P1(MCM3) and CDC46 proteins.";
RL   Nucleic Acids Res. 23:2097-2104(1995).
RN   [2]
RP   SEQUENCE OF 503-602 FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Spleen;
RX   MEDLINE=99012997; PubMed=9798653;
RA   Chu C.C., Paul W.E.;
```

```
RT   "Expressed genes in interleukin-4 treated B cells identified by cDNA
RT   representational difference analysis.";
RL   Mol. Immunol. 35:487-502(1998).
CC   -!- FUNCTION: INVOLVED IN THE CONTROL OF DNA REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26089; BAA05082.1; -.
DR   EMBL; U89402; AAC36509.1; -.
DR   PIR; S56766; S56766.
DR   MGD; MGI:103199; Mcmd4.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      457    666       MCM.
FT   NP_BIND     509    516       ATP (POTENTIAL).
FT   CONFLICT    530    530       Q -> R (IN REF. 2).
FT   CONFLICT    572    572       I -> T (IN REF. 2).
SQ   SEQUENCE   862 AA;  96736 MW;  516ACC1A3C6FB16E CRC64;

  Query Match           18.5%;  Score 856;  DB 1;  Length 862;
  Best Local Similarity   29.0%;  Pred. No. 2.8e-46;
  Matches  275;  Conservative  140;  Mismatches  331;  Indels  202;  Gaps   33;

Qy          2 ASSPAQR--RRG---------NDPLTSSPGRSSRRTDALT-------SSPGRDL--PPFE 41
              ||:|::|  |||         ::   ||| |  |  |: |        :||| ||  || :
Db          5 ASTPSRRSSRRGRVTPTQSLRSEESRSSPNRRRRGEDSSTGELLPMPTSPGADLQSPPAQ 64

Qy         42 DESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT---- 97
               :                  |     :    |||           || :    ||
Db         65 N------------------ALFSSPPQMHSLAIP-----------LDFDVSSPLTYGTP 94

Qy         98 ASRREAA--DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQ 155
              :|| |     |  ||     | ||:             |  |      |    ||   |
Db         95 SSRVEGTPRSGVRGTPVRQRPDLGS-----------ARKGLQVDLQS-----DGAAAED 137

Qy        156 MIESIENLEDLKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVD---SHGHNV------ 203
              :: | ::|    |  :  |   |::|          |: ||:   |       ||
Db        138 IVPSEQSL----GQKLVIWGTDVNVA----TCKENFQRFLQCFTDPLAKEEENVGIDITQ 189

Qy        204 --FKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYP 261
                : ::: ::       | || | : :    |   |   | |::  || |  |:    ||
Db        190 PLYMQQLGEINITGEPFLNVNCEHIKSFSKNLYRQLISYPQEVIPTFDMAVNEIFFDRYP 249

Qy        262 KYDRITNH-IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNK 320
              | |  | | ||  :    : :|:|    ::|||  ||:|   : ::|::    : |
Db        250 --DSILEHQIQVRPFNALKTKSMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQV 307

Qy        321 CNFVLGPFCQSQNQEV------KPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKV 374
              |         :   |:      :| ||  | :      :    : : : | |::||||  :
Db        308 C-------AHTTRVEIDRGRIAEPCSCVHCHTTHSMALIHNRSFFSDKQMIKLQESPEDM 360

Qy        375 AARRLPRSKDAILLA--DLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFAT 426
               | : | :   :| |  ||||    || : :|||       |         :|   |: |
Db        361 PAGQTPHT--IVLFAHNDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVSNVKSVYKT 413

Qy        427 VILANHVAKKDNKVAVGELTDED----------VKMITSLSKDQQIGEKIFASIAPSIYG 476
              |   |  | | |   |   ||:          ||::  ||:   | |:: :::|||||
Db        414 HIDVIHYRKTDAKRLHG--LDEEAEQKLFSEKRVKLLKELSRKPDIYERLASALAPSIYE 471

Qy        477 HEDIKRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIF 534
              |||||:|  | ||||  |  :  |: | | :||:|||||||:||| |:|:  :  |  :
Db        472 HEDIKKGILLQLFGGTRKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQY 531

Qy        535 TTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAME 594
              |:|:|:||| :|||| : | :|:  |: |||||:| |:| |||||||:  |: :|| ||
Db        532 TSGKGSSAVGLTAYVMKDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVME 591

Qy        595 QQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRD 654
              ||::||:|||:  | || :|:|||||  :::|  |  ||: |   ::||||:: :: |
Db        592 QQTLSIAKAGIICQLNARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLMLD 651

Qy        655 TVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYII 714
                |   |  ||  :|  : :     :||                        |  ||| ||
Db        652 PQDEAYDRRLAHHLVSLYYQSEEQVEEEF-------------------LDMAVLKDYIA 691

Qy        715 YAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLC 774
              ||   : |:|::     : : | ::||   : | :    | :||:
Db        692 YAHSTIMPRLSEEASQALIEAYVNMRKIGSSRGMVSAYPRQLESLI-------------- 737

Qy        775 DRRRRQHGHPR----------DAGELHRHTEVQRHRSMRK--------TFARYLSFRRD 815
                |   |  |           :|  |||    |     |         |     : |:
Db        738 -RLAEAHAKVRFSNKVEAIDVEEAKRLHREALKQSATDPRTGIVDISILTTGMSATSRKR 796

Qy        816 NNELLLFILKQLVAEQVT----YQR--NRFGAQQDTIEVPEKDLVDKA 857
                ||   : | ::::  |    ||:       | ||     ||: ::|
Db        797 KEELAEALRKLILSKGKTPALKYQQLFEDIRGQSDT--AITKDMFEEA 842


RESULT 11
CC54_YEAST
ID   CC54_YEAST     STANDARD;      PRT;   933 AA.
AC   P30665;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cell division control protein 54.
GN   CDC54 OR HCD21 OR YPR019W OR YP9531.13.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dalton S.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
RN   [3]
RP   SEQUENCE OF 571-646 FROM N.A.
RX   MEDLINE=93087163; PubMed=1454522;
```

```
RA   Coxon A., Maundrell K., Kearsey S.E.;
RT   "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT   early step of chromosome replication.";
RL   Nucleic Acids Res. 20:5571-5577(1992).
CC   -!- FUNCTION: REQUIRED FOR S PHASE EXECUTION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14731; AAA86310.1; -.
DR   EMBL; Z49919; CAA90164.1; -.
DR   EMBL; Z71255; CAA95015.1; -.
DR   EMBL; Z15032; CAA78750.1; -.
DR   PIR; S56050; S56050.
DR   GK; P30665; -.
DR   SGD; S0006223; CDC54.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005656; C:pre-replicative complex; IDA.
DR   GO; GO:0006267; P:pre-replicative complex formation and maint. . .; IPI.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; Cell cycle; ATP-binding.
FT   DOMAIN      518    725       MCM.
FT   NP_BIND     568    575       ATP (POTENTIAL).
SQ   SEQUENCE   933 AA;  105003 MW;  382D542AE38975E3 CRC64;

  Query Match             18.3%;  Score 848.5;  DB 1;  Length 933;
  Best Local Similarity   28.9%;  Pred. No. 9.2e-46;
  Matches  243;  Conservative  141;  Mismatches  258;  Indels  199;  Gaps   26;

Qy          5 PAQRRRGNDPLTSSPGRSSR-RTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEEL 63
              |:  :| |  :  | ||  | |: |  |: ||     :  |:  | |           |
Db         74 PSSSQRQNSDVFQSQGRQGRIRSSA--SASGRSRYHSDLRSDRALPT--------SSSSL 123

Qy         64 IGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRR----EAADGPCGTVTGSWPGLG 119
                :|  | :              :  :|    :|:: ||    :   |
Db        124 GRNGQNRVH-------------MRRNDIHTSDLSSPRRIVDFDTRSG------------ 157

Qy        120 ACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGH--SVREWVSM 177
                  :  :  : |: | | ||                 | |  : |   |::|  :
Db        158 -------VNTLDTSSSSAPPSEAS----------------EPLRIIWGTNVSIQECTT- 192

Qy        178 AGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED------------ 225
                      |:||| :          || :   :  | ||  : |  |
Db        193 ---------NFRNFLMS---------FKYKFRKILDE-REEFINNTTDEELYYIKQLNEM 233

Qy        226 ---------LAAREHVLAYFLPE--------APAELLQIFDEAALEVVLAM-------YP 261
                      | ||'::|||   |          | |:: | |:   : ::::        |
Db        234 RELGTSNLNLDAR-NLLAYKQTEDLYHQLLNYPQEVISIMDQTIKDCMVSLIVDNNLDYD 292

Qy        262 KYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKC 321
                :  |    ||  ::   : :| |     :::||   |:|   | |:| : :  : || |
Db        293 LDEIETKFYKVRPYNVGSCKGMRELNPNDIDKLINLKGLVLRSTPVIPDMKVAFFKCNVC 352

Qy        322 NFVLGPFCQSQNQEV------KPGSCP--ECQSAGPFEVNMEETIYQNYQRIRIQESPGK 373
                :  :         |:       :|  |   :|       :      : : | |::||:|
Db        353 DHTMA-------VEIDRGVIQEPARCERIDCNEPNSMSLIHNRCSFADKQVIKLQETPDF 405

Qy        374 VAARRLPRSKDAILLADLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATV 427
              |   : | |    :  :|||| ||| ||:|| :       |:    | :     :: |

Db        406 VPDGQTPHSISLCVYDELVDSCRAGDRIEVTGTFRSIPIRANSRQRVLKS-----LYKTY 460

Qy        428 ILANHVAK----------------KDNKV------AVGELTDEDVKMITSLSKDQQIGE 464
                :  || |                    |||       | ::||:|:  |  ::  : :
Db        461 VDVVHVKKVSDKRLDVDTSTIEQELMQNKVDHNEVEEVRQITDQDLAKIREVAAREDLYS 520

Qy        465 KIFASIAPSIYGHEDIKRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGTAKSQFL 522
                 :  ||||||| ||:|:|  | |||| |    ||::  |||||:|||||| |:||| |
Db        521 LLARSIAPSIYELEDVKKGILLQLFGGTNKTFTKGGRY--RGDINILLCGDPSTSKSQIL 578

Qy        523 KYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMN 582
                :|: |:: | ::|:|:|:|:||| :|||: |   :::  ||:||||||:| ||| |||||||:
Db        579 QYVHKITPRGVYTSGKGSSAVGLTAYITRDVDTKQLVLESGALVLSDGGVCCIDEFDKMS 638

Qy        583 DQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPI 642
              |  |: :|| ||||:|||:||||:|:| || :::|:|||| ||:|:|  :||:||  |:
Db        639 DSTRSVLHEVMEQQTISIAKAGIITTLNARSSILASANPIGSRYNPNLPVTENIDLPPPL 698

Qy        643 ISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVE 702
                :||||:: :| | ||   |  ||: :   ::   | :  ::                :
Db        699 LSRFDLVYLVLDKVDEKNDRELAKHLTNLYLEDKPEHISQD----------------D 740

Qy        703 PLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGS----IPITVRHIES 758
                || | |  || |||| :|| : :  : :: : |  :||    : |    |  | | :||
Db        741 VLPVEFLTMYISYAKEHIHPIITEAAKTELVRAYVGMRKMGDDSRSDEKRITATTRQLES 800

Qy        759 M 759
              |
Db        801 M 801
```

```
RESULT 12
CC21_SCHPO
ID   CC21_SCHPO     STANDARD;      PRT;   931 AA.
AC   P29458; Q9P7K4; Q9USM0;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cdc21 protein.
GN   CDC21 OR SPCC16A11.17 OR SPCC24B10.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93087163; PubMed=1454522;
RA   Coxon A., Maundrell K., Kearsey S.E.;
RT   "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT   early step of chromosome replication.";
RL   Nucleic Acids Res. 20:5571-5577(1992).
RN   [2]
RP   REVISIONS TO C-TERMINUS.
RA   Kearsey S.E.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
```

```
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: REQUIRED FOR S PHASE EXECUTION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58824; CAA41628.1; -.
DR   EMBL; AL109957; CAB53089.1; -.
DR   EMBL; AL157991; CAB76210.1; -.
DR   GeneDB_SPombe; SPCC16A11.17; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      493    702       MCM.
FT   NP_BIND     545    552       ATP (POTENTIAL).
FT   CONFLICT    872    891       MISSING (IN REF. 3).
SQ   SEQUENCE   931 AA;  103728 MW;  BCA9B045FC62811D CRC64;

  Query Match           18.2%;  Score 843;  DB 1;  Length 931;
  Best Local Similarity   28.5%;  Pred. No. 2e-45;
  Matches  236;  Conservative  164;  Mismatches  289;  Indels  140;  Gaps   28;

Qy          2 ASSPAQRRRGNDPL---TSSPGR---SSRRTDALT-----------SSPGRDLPPFEDES 44
              |:| ::    : ||    ||||    :: || | |           |||| ::|  :
Db         19 ANSSSREAVDSSPLFFPASSPGSTRLTTPRTTARTPLASSPLLFESSSPGPNIP--QSSR 76

Qy         45 EGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA 104
                ||     |         :    :|  |:    | : :          |:  | |||    ~
Db         77 SHLLSQRNDL--------FLDSSSQRTPRSTRRGDIHSS---------VQMSTPSRRREV 119

Qy        105 DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLE 164
              |          ||:              | |:|  | :            :: :    :
Db        120 D-------PQRPGVS------------TPSSLLFSGSDALTFSQAHPSSEVADDTVRVI 159

Qy        165 DLKGHSVREWVS-----MAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESL 219
                  |::| ::     : | : :    ::| |    |:    |: | : :|     | |
Db        160 WGTNVSIQESIASFRGFLRGFKKKYRPEYRNELMPPPDAE-QLVYIEALRNMRIMGLEIL 218

Qy        220 VVNYEDLAAREHV-----LAYFLPEAPAELLQIFDEAALEVVLAMY---PKYDRITNHIH 271
                :: :||   :|      | : |   | |:: | |:   :|:| :    |  | : | |
Db        219 NLDVQDL---KHYPPTKKLYHQLYSYPQEIIPIMDQTIKDVMLDLLGTNPPED-VLNDIE 274

Qy        272 VRISHL-PL-VEELRSLRQLH---LNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLG 326
              ::|  : |  :|:  ::| |:    :::||   |:|  || |:| :    : |: |
Db        275 LKIYKIRPFNLEKCINMRDLNPGDIDKLISIKGLVLRCTPVIPDMKQAFFRCSVCG---- 330

Qy        327 PFCQS----QNQEVKPGSCPE--CQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLP 380
                | :    : :  :|  ||   | :    ::    : : : | |::||:|  |   : |
Db        331 -HCVTVEIDRGRIAEPIKCPREVCGATNAMQLIHNRSEFADKQVIKLQETPDVVPDGQTP 389

Qy        381 RSKDAILLADLVDSCNAGDEIELTGIYH---NNYDGSLNTANGFPVFATVILANHVAKKD 437
                |    : :||||  ||| ||:|||:       :  : |     :| | :   |: |:|
Db        390 HSVSLCVYDELVDSARAGDRIEVTGIFRCVPVRLNPRMRTVKS--LFKTYVDVVHIKKQD 447

Qy        438 NK----------VA---------VGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYG 476
                :           :|         | :::||:|: |  :||  | : :  |:|||||
Db        448 KRRLGTDPSTLESDIAEDAALQIDEVRKISDEEVEKIQQVSKRDDIYDILSRSLAPSIYE 507

Qy        477 HEDIKRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIF 534
              :|:|:|  | ||||  |  : |   : |||||:|:|||| |:||| |||: |:: | ::
Db        508 MDDVKKGLLLQLFGGTNKSFHKGASPRYRGDINILMCGDPSTSKSQILKYVHKIAPRGVY 567

Qy        535 TTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAME 594
              |:|:|:||| :|||: |   :::  ||:|||||:| |:| |||||||:|  |: :|| ||
Db        568 TSGKGSSAVGLTAYITRDQDTKQLVLESGALVLSDGGICCIDEFDKMSDATRSILHEVME 627

Qy        595 QQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRD 654
              ||:::::||||:|:| || :::|:|||| :|:| |  ::|:||   ::||||:: :: |
Db        628 QQTVTVAKAGIITTLNARTSILASANPIGSKYNPDLPVTKNIDLPPTLLSRFDLVYLILD 687

Qy        655 TVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYII 714
              ||   |  ||  :|  ::   | :  :                    :|    | |  ||
Db        688 RVDETLDRKLANHIVSMYMEDTPEHATD----------------MEVFSVEFLTSYIT 729

Qy        715 YAKERVHPKLNQMDQDKVAKMYSDLRK--ESMATGS--IPITVRHIESM 759
              ||:  ::| :::    ::   |  :||  | :      |  | | :|||
Db        730 YARNNINPVISEEAAKELVNAYVGMRKLGEDVRASEKRITATTRQLESM 778


RESULT 13
MCM6_MOUSE
ID   MCM6_MOUSE     STANDARD;      PRT;   821 AA.
AC   P97311;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   DNA replication licensing factor MCM6 (Mis5 homolog).
GN   MCM6 OR MCMD6 OR MIS5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=97224213; PubMed=9077461;
RA   Kimura H., Ohtomo T., Yamaguchi M., Ishii A., Sugimoto K.;
RT   "Mouse MCM proteins: complex formation and transportation to the
RT   nucleus.";
RL   Genes Cells 1:977-993(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE CONTROL OF A SINGLE ROUND OF DNA
CC       REPLICATION DURING S PHASE. BINDS TO CHROMATIN DURING G1 AND
CC       DETACH FROM IT DURING S PHASE AS IF IT LICENSES THE CHROMATIN TO
CC       REPLICATE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86726; BAA13159.1; -.
DR   MGD; MGI:1298227; Mcmd6.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
```

```
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding; Cell cycle.
FT   DOMAIN      346    553       MCM.
FT   NP_BIND     396    403       ATP (POTENTIAL).
SQ   SEQUENCE   821 AA;  92867 MW;  79E753D2DA218CBF CRC64;

  Query Match             18.1%;  Score 840;  DB 1;  Length 821;
  Best Local Similarity   27.3%;  Pred. No. 2.6e-45;
  Matches  234;  Conservative  154;  Mismatches  286;  Indels  184;  Gaps   21;

Qy        137 LPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHV 196
              | |:|  |||::  |  ::: |  : |                          | :||
Db          3 LAAAAEPGAGSQHPEVRDEVAEKCQKL-----------------------FLDFLEEFQ 38

Qy        197 DSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEA-------PAELLQIFD 249
              | |   : :   :: :  | :|||:: ||      |:  : |           |: |
Db         39 GSDGEIKYLQFAEELIRPERNTLVVSFADLEQFNQQLSTTIQEEFYRVYPYLCRALKTFV 98

Qy        250 EAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLP 309
              :   |:  |           :|    ||   ::| |     :  | | || |     | |
Db         99 KDRKEIPFA---------KDFYVAFQDLPTRHKIRELTSSRIGLLTRISGQVVRTHPVHP 149

Qy        310 QLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC--PECQSAGPFEVNMEETIYQNYQRIRI 367
              :|    : |  | |:    : | :  :|  |  | | :   | ::  :: : ::|::||
Db        150 ELVSGTFLCLDCQTVIKD-VEQQFKYTQPNICRNPVCANRKRFLLDTNKSRFVDFQKVRI 208

Qy        368 QESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIY--------------HNNYDG 413
              ||:  ::    :||| : || |: |:|  |||  : ||                :
Db        209 QETQAELPRGSIPRSLEVILRAEAVESAQAGDRCDFTGALIVVPDVSKLSTPGARAETNS 268

Qy        414 SLNTANGFPV---------------FATVILANHVAKKDNKVAVGELTDEDV-------- 450
              :: |:|:                 :  | || |||  : :    || ||:
Db        269 RVSGADGYETEGIRGLRALGVRDLSYRLVFLACHVAPTNPRFGGKELRDEEQTAESIKNQ 328

Qy        451 ------KMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRG 504
                   : :  :|:|: :   :  |: |:|:|::::||| | |||| || |    :||
Db        329 MTVKEWEKVFEMSQDKNLYHNLCTSLFPTIHGNDEVKRGVLLMLFGGVPKTTGEGTSLRG 388

Qy        505 DINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGA 564
              |||| : ||| |||||||:::: | ||::|:|: :||  :|| | |    | |: :||||
Db        389 DINVCIVGDPSTAKSQFLKHVDEFSPRAVYTSGKASSAAGLTAAVVRDEESHEFVIEAGA 448

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGG 624
              |:||| ||| |||||||: :|: :|||||||:|||:|||:  :| || :::|||||: |
Db        449 LMLADNGVCCIDEFDKMDMRDQVAIHEAMEQQTISITKAGVKATLNARTSILAAANPVSG 508

Qy        625 RYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEG 684
              || | :  :|:::|: ||:||||:  :: |  : | |  :|| :|  |
Db        509 HYDRSKSLKQNINLSAPIMSRFDLFFILVDECNEVTDYAIARRIVDLH------------ 556

Qy        685 LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK--- 741
              |  | ::   | ::      :::|:::|::   ||:::  :| : : |  ||:
Db        557 ----SRIEESIDRVYSLDD-----IRRYLLFARQ-FKPKISKESEDFIVEQYKRLRQRDG 606

Qy        742 ESMATGSIPITVRHIESMSH--GGGPRAHPSAGLCDRRRRQH----------------- 781
              :   |  |||| :|||        | |     ||  : :|
Db        607 SGVTKSSWRITVRQLESMIRLSESMARMH----CCDEVQPKHVKEAFRLLNKSIIRVETP 662

Qy        782 ----------------------GHPRDAGELHRHT--------EVQRHRSMRKTFARYLS 811
                                   ||      ::|          |     |:|  || |
Db        663 DVNLDQEEEIQMETDEGQGGVNGHADSPAPVNRFNGSSEDASQETVSKPSLRLGFAEYC- 721

Qy        812 FRRDNNELLLFILKQ---------LVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQIN- 861
              | :| ::| : |                :| | :      :: |:    |::|::|   |
Db        722 --RISNLIVLHLRKMEEEEDESALKRSELVNWYLKEIESEIDS----EEELINKKTIIEK 775

Qy        862 -IHNLSAFYDSELFRMNK 878
              :| |:  ||  |  : :
Db        776 VVHRLT-HYDHVLIELTQ 792


RESULT 14
```

```
MCM6_HUMAN
ID   MCM6_HUMAN     STANDARD;      PRT;   821 AA.
AC   Q14566; Q13504; Q99859;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA replication licensing factor MCM6 (P105MCM).
GN   MCM6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97431351; PubMed=9286856;
RA   Tsuruga H., Yabuta N., Hosoya S., Tamura K., Endo Y., Nojima H.;
RT   "HsMCM6: a new member of the human MCM/P1 family encodes a protein
RT   homologous to fission yeast Mis5.";
RL   Genes Cells 2:381-399(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98184833; PubMed=9516426;
RA   Holthoff H.P., Baack M., Richter A., Ritzi M., Knippers R.;
RT   "Human protein MCM6 on HeLa cell chromatin.";
RL   J. Biol. Chem. 273:7320-7325(1998).
RN   [3]
RP   SEQUENCE OF 640-821 FROM N.A.
RX   MEDLINE=97131582; PubMed=8977093;
RA   Harvey C.B., Wang Y., Darmoul D., Phillips A., Mantei N.,
RA   Swallow D.M.;
RT   "Characterisation of a human homologue of a yeast cell division cycle
RT   gene, MCM6, located adjacent to the 5' end of the lactase gene on
RT   chromosome 2q21.";
RL   FEBS Lett. 398:135-135(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE CONTROL OF A SINGLE ROUND OF DNA
CC       REPLICATION DURING S PHASE. BINDS TO CHROMATIN DURING G1 AND
CC       DETACH FROM IT DURING S PHASE AS IF IT LICENSES THE CHROMATIN TO
CC       REPLICATE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D84557; BAA12699.1; -.
DR   EMBL; U46838; AAC50766.1; -.
DR   EMBL; U67284; AAB48165.1; -.
DR   EMBL; U67281; AAB48165.1; JOINED.
DR   EMBL; U67282; AAB48165.1; JOINED.
DR   EMBL; U67283; AAB48165.1; JOINED.
DR   Genew; HGNC:6949; MCM6.
DR   GK; Q14566; -.
DR   MIM; 601806; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0005524; F:ATP binding activity; NAS.
DR   GO; GO:0006260; P:DNA replication; NAS.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding; Cell cycle; Polymorphism.
FT   DOMAIN      346    553       MCM.
FT   NP_BIND     396    403       ATP (POTENTIAL).
FT   VARIANT      35     35       E -> V (IN dbSNP:3087355).
FT                                /FTId=VAR_014816.
```

```
FT   CONFLICT    377    387       PKTTGEGTSLR -> SKDNRRRDLSS (IN REF. 2).
FT   CONFLICT    495    495       A -> T (IN REF. 2).
FT   CONFLICT    738    738       MISSING (IN REF. 3).
FT   CONFLICT    790    790       L -> P (IN REF. 2).
SQ   SEQUENCE   821 AA;  92889 MW;  F94968EB25A3E501 CRC64;

  Query Match             18.0%;  Score 835;  DB 1;  Length 821;
  Best Local Similarity   27.5%;  Pred. No. 5.4e-45;
  Matches  236;  Conservative  155;  Mismatches  284;  Indels  182;  Gaps   22;

Qy        137 LPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHV 196
              | |:|  |||::  |  ::: |  : |                          | :||
Db          3 LAAAAEPGAGSQHLEVRDEVAEKCQKL------------------------FLDFLEEFQ  38

Qy        197 DSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEA-------PAELLQIFD 249
              | |   : :   :: :  | :|||:: ||      |:  : |             |: |
Db         39 SSDGEIKYLQLAEELIRPERNTLVVSFVDLEQFNQQLSTTIQEEFYRVYPYLCRALKTFV  98

Qy        250 EAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLP 309
              :  |: ||           :|    ||   ::| |    :  | | || |     | |
Db         99 KDRKEIPLA---------KDFYVAFQDLPTRHKIRELTSSRIGLLTRISGQVVRTHPVHP 149

Qy        310 QLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC--PECQSAGPFEVNMEETIYQNYQRIRI 367
              :|    : |  |  |:    : | :  :|  |  | | :   | ::  :: : ::|::||
Db        150 ELVSGTFLCLDCQTVIRD-VEQQFKYTQPNICRNPVCANRRRFLLDTNKSRFVDFQKVRI 208

Qy        368 QESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIY--------------HNNYDG 413
              ||:  ::    :||| : || |: |:|  |||: : ||                 :
Db        209 QETQAELPRGSIPRSLEVILRAEAVESAQAGDKCDFTGTLIVVPDVSKLSTPGARAETNS 268

Qy        414 SLNTANGFPV---------------FATVILANHVAKKDNKVAVGELTDEDV-------- 450
              ::  :|:                 :  | ||  ||  : :    || ||:
Db        269 RVSGVDGYETEGIRGLRALGVRDLSYRLVFLACCVAPTNPRFGGKELRDEEQTAESIKNQ 328

Qy        451 ------KMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRG 504
                   : :  :|:|: :   :  |: |:|:|:::::|||  | |||| ||  |    :||
Db        329 MTVKEWEKVFEMSQDKNLYHNLCTSLFPTIHGNDEVKRGVLLMLFGGVPKTTGEGTSLRG 388

Qy        505 DINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGA 564
              |||| : ||| |||||||::|: | ||::|:|: :||  :|| | |    | |: :||||
Db        389 DINVCIVGDPSTAKSQFLKHVEEFSPRAVYTSGKASSAAGLTAAVVRDEESHEFVIEAGA 448

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGG 624
              |:||| ||| |||||||: :|: :|||||||:|||:|||:  :| || :::|||||| |
Db        449 LMLADNGVCCIDEFDKMDVRDQVAIHEAMEQQTISITKAGVKATLNARTSILAAANPISG 508

Qy        625 RYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEG 684
              || | :  :|::|: ||:||||:  :: |  : | |  :|| :|  |
Db        509 HYDRSKSLKQNINLSAPIMSRFDLFFILVDECNEVTDYAIARRIVDLH------------ 556

Qy        685 LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK--- 741
                | | ::   | ::       :::|:::|::   ||::: :| : : |  ||:
Db        557 ----SRIEESIDRVYSLDD-----IRRYLLFARQ-FKPKISKESEDFIVEQYKHLRQRDG 606

Qy        742 ESMATGSIPITVRHIESMSH--GGGPRAHPSAGLCDRRRRQH----------------- 781
              :  | |||| :|||       | |     ||  : :|
Db        607 SGVTKSSWRITVRQLESMIRLSEAMARMH----CCDEVQPKHVKEAFRLLNKSIIRVETP 662

Qy        782 ---------------------GH---PRDAGELHRHTEVQRHRSMRKTFAR--YLSFRR 814
                                  ||   |     :: : |     | |   |  :  : |
Db        663 DVNLDQEEEIQMEVDEGAGGINGHADSPAPVNGINGYNEDINQESAPKASLRLGFSEYCR 722

Qy        815 DNNELLLFI-----------LKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQIN-- 861
              :| ::| :           ||:  :| | :      :: |:    |::|::| | |
Db        723 ISNLIVLHLRKVEEEEDESALKR--SELVNWYLKEIESEIDS----EEELINKKRIIEKV 776

Qy        862 IHNLSAFYDSELFRMNK 878
              || |:  ||  |  : :
Db        777 IHRLT-HYDHVLIELTQ 792


RESULT 15
MCM5_XENLA
ID   MCM5_XENLA     STANDARD;      PRT;   735 AA.
AC   P55862;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA replication licensing factor MCM5 (CDC46 homolog) (X.CDC46).
GN   MCM5 OR CDC46.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=97357318; PubMed=9214647;
RA   Kubota Y., Mimura S., Nishimoto S., Masuda T., Nojima H., Takisawa H.;
RT   "Licensing of DNA replication by a multi-protein complex of MCM/P1
RT   proteins in Xenopus eggs.";
RL   EMBO J. 16:3320-3331(1997).
RN   [2]
RP   SEQUENCE OF 20-735 FROM N.A.
RC   TISSUE=Oocyte;
RX   MEDLINE=97074651; PubMed=8917078;
RA   Miyake S., Saito I., Kobayashi H., Yamashita S.;
RT   "Identification of two Xenopus laevis genes, xMCM2 and xCDC46, with
RT   sequence homology to MCM genes involved in DNA replication.";
RL   Gene 175:71-75(1996).
CC   -!- FUNCTION: INVOLVED IN EARLY STEP OF CHROMOSOME REPLICATION.
CC   -!- SUBUNIT: ASSOCIATED AND INTERACTS WITH MCM2.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MCM FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U44048; AAC60224.1; -.
DR   EMBL; D63920; BAA09949.1; -.
DR   PIR; T47224; PC4225.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   DNA replication; ATP-binding.
FT   DOMAIN      332    538       MCM.
FT   NP_BIND     382    389       ATP (POTENTIAL).
SQ   SEQUENCE   735 AA;  82435 MW;  702BA90C2F510720 CRC64;

  Query Match             17.8%;  Score 825.5;  DB 1;  Length 735;
  Best Local Similarity   30.9%;  Pred. No. 1.8e-44;
  Matches  231;  Conservative  127;  Mismatches  302;  Indels   87;  Gaps   18;

Qy        183 EIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRES----LVVNYEDLAAREHVLAYFLP 238
              ::  ||: ||| :        |  :  |  | :       : |  ||||: :  || :|
Db         31 QLKKRFREFLRQYRIGTDRTGFTFKYRDELKRHYNLGEYWIEVEMEDLASFDEDLADYLY  90

Qy        239 EAPAELLQIFDEAALEV---VLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLI 295
              : | | ||: :||| ||   |    |  :     | | :         :|||:   :: |:
Db         91 KQPTEHLQLLEEAAQEVADEVTRPRPAGEETIQEIQVMLRSDANPANIRSLKSEQMSHLV 150

Qy        296 RTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPG--------SC------ 341
              :  |:: : | |  : : :   |  |   :|      |  |:||         |
Db        151 KIPGIIIAATAVRAKATKISIQCRSCRNTIG------NIAVRPGLEGYAMPRKCNTEQAG 204
```

```
Qy        342 -PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDE 400
Db        205 RPNC-PLDPYFIIPDKCKCVDFQTLKLQESPDAVPHGELPRHMQLYCDRYLCDKVVPGNR 263

Qy        401 IELTGIYHNNYDGSLNTANGFPV-------FATVILANHVAKKDNKVAVGELTDEDVKMI 453
Db        264 VTIMGIYSIRKSGKTSTKGRDRVGVGIRSSYIRVVGIQVDTEGTGRSAAGAITPQEEEEF 323

Qy        454 TSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKN-PGGKHKVRGDINVLLCG 512
Db        324 RRLAAKPDIYETVAKSIAPSIYGSSDIKKAIACLLFGGSRKRLPDGLTR-RGDVNLLMLG 382

Qy        513 DPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGV 572
Db        383 DPGTAKSQLLKFVERCSPIGVYTSGKGSSAAGLTASVMRDPVSRNFIMEGGAMVLADGGV 442

Qy        573 CLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTF 632
Db        443 VCIDEFDKMREDDRVAIHEAMEQQTISIAKAGITTTLNSRCSVLAAANSVYGRWDDT-KG 501

Qy        633 SENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAE 692
Db        502 EENIDFMPTILSRFDMIFIVKDEHNEQRDMTLAKHVMNVHLSARTQSSSVEG-------- 553

Qy        693 PAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLR-------KESMA 745
Db        554 ----------EVDLNTLKKYIAYCRAKCGPRLSAEAAEKLKNRYILMRSGAREHERETEK 603

Qy        746 TGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQR-HRSMRK 804
Db        604 RSSIPITVRQLEAI-------VRISESLGKMKLQPFATETDVEEALRLFQVSTLDAAMSG 656

Qy        805 TFARYLSF-RRDNNELLLFILKQLVAEQVTYQRNRF--GAQQDTIEVPEKDLVDKARQIN 861
Db        657 SLSGVEGFTTQEDQEMLSRIEKQM--------KKRFAIGSQVSEHSIIQDFLKQKYPEHA 708

Qy        862 IHNLSAFYDSELFRMNKFSHDLKRKMI 888
Db        709 IHKVL----SLMMRRGEIQHRLQRKVL 731
```

Search completed: September 12, 2003, 13:12:49
Job time : 21 secs

OM protein - protein search, using sw model

Run on: September 12, 2003, 13:10:05 ; Search time 54 Seconds
(without alignments)
4262.647 Million cell updates/sec

Title: CAA47749
Perfect score: 4630
Sequence: 1 MASSPAQRRRGNDPLTSSPG.....LFRMNKFSHDLKRKMILQQF 892

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3955	85.4	904	11	Q8C2R0	Q8c2r0 mus musculu
2	3942	85.1	904	11	O08971	O08971 mus musculu
3	3634.5	78.5	811	4	Q8N2V1	Q8n2v1 homo sapien
4	3551.5	76.7	886	13	O42588	O42588 xenopus lae
5	3192.5	69.0	880	13	Q8JGR5	Q8jgr5 brachydanio
6	2254	48.7	881	5	Q9XXI9	Q9xxi9 caenorhabdi
7	2132	46.0	871	5	Q9BI22	Q9bi22 dugesia jap
8	1945.5	42.0	936	10	Q9LPD9	Q9lpd9 arabidopsis
9	1887.5	40.8	865	10	Q8H0G9	Q8h0g9 nicotiana t
10	1864.5	40.3	889	3	O42723	O42723 emericella
11	1604.5	34.7	365	4	Q9BWF4	Q9bwf4 homo sapien
12	1525	32.9	780	5	Q8SS42	Q8ss42 encephalito
13	1277	27.6	883	5	Q9U446	Q9u446 entamoeba h
14	1267	27.4	971	5	Q8ILR7	Q8ilr7 plasmodium
15	1262	27.3	971	5	Q9GR05	Q9gr05 plasmodium
16	1253	27.1	491	10	Q96275	Q96275 arabidopsis
17	913	19.7	686	17	Q9UXG1	Q9uxg1 sulfolobus
18	909.5	19.6	680	17	Q8ZY88	Q8zy88 pyrobaculum
19	909.5	19.6	699	17	Q9YFR1	Q9yfr1 aeropyrum p
20	872	18.8	863	4	Q8NEH1	Q8neh1 homo sapien
21	858	18.5	720	10	Q9SIV8	Q9siv8 arabidopsis
22	856	18.5	849	10	Q8RYQ2	Q8ryq2 oryza sativ
23	855	18.5	776	10	O65400	O65400 arabidopsis
24	853	18.4	776	10	Q9FL33	Q9fl33 arabidopsis
25	853	18.4	862	11	Q9D077	Q9d077 mus musculu
26	852	18.4	862	11	Q8C1Z0	Q8c1z0 mus musculu
27	846	18.3	862	11	Q921D5	Q921d5 mus musculu
28	843.5	18.2	858	13	O42589	O42589 xenopus lae
29	839.5	18.1	768	10	Q9SX03	Q9sx03 zea mays (m
30	839.5	18.1	768	10	Q9SX04	Q9sx04 zea mays (m
31	825.5	17.8	701	17	Q8TSW4	Q8tsw4 methanosarc
32	823.5	17.8	666	17	O27798	O27798 methanobact
33	821.5	17.7	824	13	O73710	O73710 xenopus lae
34	819	17.7	819	5	Q9XYU1	Q9xyu1 drosophila
35	812	17.5	819	5	P91675	P91675 drosophila
36	810.5	17.5	701	17	Q8PVX1	Q8pvx1 methanosarc
37	807.5	17.4	796	13	O42590	O42590 xenopus lae
38	800	17.3	733	5	Q9VGW6	Q9vgw6 drosophila
39	797	17.2	823	5	Q95XQ8	Q95xq8 caenorhabdi
40	796.5	17.2	727	10	O80786	O80786 arabidopsis
41	788	17.0	720	13	O42592	O42592 xenopus lae
42	788	17.0	726	5	Q8SRX5	Q8srx5 encephalito
43	786	17.0	710	10	Q8H1A2	Q8h1a2 pisum sativ
44	784	16.9	586	17	O29733	O29733 archaeoglob
45	784	16.9	812	5	Q9XVR7	Q9xvr7 caenorhabdi

ALIGNMENTS

RESULT 1
Q8C2R0
ID Q8C2R0 PRELIMINARY; PRT; 904 AA.
AC Q8C2R0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mini chromosome maintenance deficient 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK088156; BAC40178.1; -.
SQ SEQUENCE 904 AA; 102077 MW; 7B61C13DAD1CAC58 CRC64;

Query Match 85.4%; Score 3955; DB 11; Length 904;
Best Local Similarity 86.2%; Pred. No. 3.1e-276;
Matches 788; Conservative 23; Mismatches 61; Indels 42; Gaps 7;

```
Qy          2 ASSPA-QRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||| |||| :|||||||||||||| |||||||||||||||||||||||||:||||||
Db         10 ASSPARQRRRISDPLTSSPGRSSRRADALTSSPGRDLPPFEDESEGLLGTEGPMEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              |||||||||||||| ||||| |||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRPIPELDVYEAEGLALDDEDVEELTASQREAAERTMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTAMRRTRSALPASAASG------------AGTEDGEEDEQMIESIE 161
                 |||           |  |  |  |:                 ||||||||:||||||
Db        127 ---GLG-----------RMRRGLLYDSSEEDEERPARKRRHVERATEDGEEDEEMIESIE 172
```

```
Qy        162 NLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 NLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVV 232

Qy        222 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE 292

Qy        282 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC 341
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        293 ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCSKCNFVLGPFCQSQNQEVKPGSC 352

Qy        342 PECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEI 401
              ||||||||||:|||||||||||||||||||||||| ||||||||||||||||  ||||
Db        353 PECQSAGPFEINMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEI 412

Qy        402 ELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQ 461
              ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db        413 ELTGIYHNNYDGSLNTANGFPVFATIILANHVAKKDNKVAVGELTDEDVKMITGLSKDQQ 472

Qy        462 IGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQF 521
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        473 IGEKIFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQF 532

Qy        522 LKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKM 581
              |||||||||||||||||||| :|||||||||||||||||||||||||||||||||||||||
Db        533 LKYIEKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKM 592

Qy        582 NDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEP 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        593 NDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEP 652

Qy        642 IISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGV 701
              ||||||:|||||||||||||||||||||||||||||||::||| ||   ||||||||||
Db        653 IISRFDVLCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKKDEGLTNGGTLEPAMPNTYGV 712

Qy        702 EPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSH 761
              |||||||||||||||||| ||||||||||:|||||||||||||||||||||||||||||
Db        713 EPLPQEVLKKYIIYAKERVRPKLNQMDQDKVARMYSDLRKESMATGSIPITVRHIESMIR 772

Qy        762 GGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNE 818
                    ||    | |       :      :    : |:    ||||||||||||||||:
Db        773 MA--EAHARMHLRDYVMEDDVNMAIRVMMESFIDTQKFSVMRSMRKTFARYLSFRRDNND 830

Qy        819 LLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNK 878
              ||||||||||||||||||||||||||||:|||||:||||||||||||||||||:||: ||
Db        831 LLLFILKQLVAEQVTYQRNRFGAQQDTIEIPEKDLMDKARQINIHNLSAFYDSDLFKFNK 890

Qy        879 FSHDLKRKMILQQF 892
              || |||||:|||||
Db        891 FSRDLKRKLILQQF 904


RESULT 2
O08971
ID   O08971      PRELIMINARY;      PRT;   904 AA.
AC   O08971;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   BM28 homolog.
GN   MCMD2 OR CDCL1 HOMOLOG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=98037564; PubMed=9371513;
RA   Spanjaard R.A., Lee P.J., Sarkar S., Goedegebuure P.S., Eberlein T.J.;
RT   "Clone 10d/BM28 (CDCL1), an early S-phase protein, is an important
RT   growth regulator of melanoma.";
RL   Cancer Res. 57:5122-5128(1997).
DR   EMBL; AF004105; AAC16250.1; -.
DR   MGD; MGI:105380; Mcmd2.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   904 AA;  101963 MW;  8B982C67FABBF0D0 CRC64;

  Query Match             85.1%;  Score 3942;  DB 11;  Length 904;
  Best Local Similarity   86.7%;  Pred. No. 2.7e-275;
  Matches  784;  Conservative   25;  Mismatches   73;  Indels   22;  Gaps    7;

Qy          2 ASSPA-QRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||| |||| :|||||||||||||| |||||||||||||||||||||||||:||||||
Db         10 ASSPARQRRRISDPLTSSPGRSSRRADALTSSPGRDLPPFEDESEGLLGTEGPMEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
               |||||||||||| ||||| ||||||||||||||||:||||       |   |
Db         70 GELIGDGMERDYRPIPELDVYEAEGLALDDEDVEELTASQREAAERTMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTAMRRTRSALPASAASGA--GTEDGEEDEQMIESIENLEDLKGHSV 171
                 |||    |    :        ||         ||||||||:|||||||||||||||
Db        127 ---GLGRMRRGPLYDSSEEDEER-PARKRRHVERATEDGEEDEEMIESIENLEDLKGHSV 182

Qy        172 REWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREH 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 REWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREH 242

Qy        232 VLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHL 291
              ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        243 VLAYFLPEAPAELLQIFDEAALEVVLAMHPKYDRITNHIHVRISHLPLVEELRSLRQLHL 302

Qy        292 NQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFE 351
              ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db        303 NQLIRTSGVVTSCTGVLPQLSMVKYNCSKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFE 362

Qy        352 VNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNY 411
              :|||||||||||||||||||||||| ||||||||||||||  ||||||||||||||
Db        363 INMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTGIYHNNY 422

Qy        412 DGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIA 471
              |||||||||||||:|||||||||||||||||||||||||||||| ||||||||||||||
Db        423 DGSLNTANGFPVFATIILANHVAKKDNKVAVGELTDEDVKMITGLSKDQQIGEKIFASIA 482

Qy        472 PSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSR 531
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 PSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSR 542

Qy        532 AIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHE 591
              |||||||||||| :|||||||||||||||||||||||||||||||||||||||||||||||
Db        543 AIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHE 602

Qy        592 AMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCV 651
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        603 AMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDVLCV 662

Qy        652 VRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKK 711
              ||||||||||||||||||||||||||::||| ||   ||||||||||||||||||||
Db        663 VRDTVDPVQDEMLARFVVGSHVRHHPSNKKDEGLTNGGTLEPAMPNTYGVEPLPQEVLKK 722

Qy        712 YIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSA 771
              ||||||||| |||||||||||:|||||||||||||||||||||||||        ||
Db        723 YIIYAKERVRPKLNQMDQDKVARMYSDLRKESMATGSIPITVRHIESMIRMA--EAHARM 780

Qy        772 GLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLV 828
              | |       :      :    : |:    ||||||||||:|||||||:||||||||||
Db        781 HLRDYVMEDDVNMAIRVMMESFIDTQKFSVMRSMRKTFARYISFRRDNNDLLLFILKQLV 840

Qy        829 AEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMI 888
```

```
                    ||||||||||||||||||:|||||:||||||||||||||||:||: |||| |||||:|
Db        841 AEQVTYQRNRFGAQQDTIEIPEKDLMDKARQINIHNLSAFYDSDLFKFNKFSRDLKRKLI 900

Qy        889 LQQF 892
              ||||
Db        901 LQQF 904


RESULT 3
Q8N2V1
ID   Q8N2V1      PRELIMINARY;      PRT;   811 AA.
AC   Q8N2V1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC030131; AAH30131.1; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   811 AA;  91850 MW;  82556224C09FEFD7 CRC64;

  Query Match           78.5%;  Score 3634.5;  DB 4;  Length 811;
  Best Local Similarity   88.5%;  Pred. No. 3.4e-253;
  Matches  731;  Conservative    8;  Mismatches   54;  Indels   33;  Gaps    6;

Qy         85 GLALDDEDVEELTASRREAA-------DGPCGTVTGSWPGLGACAVGSCMTA-------- 129
              |  ||||||||||||:||||       |   |       |||    |    :
Db          1 GTRLDDEDVEELTASQREAAERAMRQRDREAGR------GLGRMRRGLLYDSDEEDEERP 54

Qy        130 MRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFK 189
               |: |    |       ||||||||:|||||||||||||||||||||||||||||||||||
Db         55 ARKRRQVERA-------TEDGEEDEEMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFK 107

Qy        190 NFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFD 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        108 NFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFD 167

Qy        250 EAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLP 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 EAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLP 227

Qy        310 QLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQE 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 QLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQE 287

Qy        370 SPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVIL 429
              ||||||| ||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        288 SPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTGIYHNNYDGSLNTANGFPVFATVIL 347

Qy        430 ANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALF 489
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 ANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGLALALF 407

Qy        490 GGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYV 549
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 GGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVGLTAYV 467

Qy        550 QRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSL 609
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 QRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSL 527

Qy        610 QARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVV 669
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 QARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVV 587

Qy        670 GSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQ 729
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQ 647

Qy        730 DKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE 789
              |||||||||||||||||||||||||||||||       ||    | |       :
Db        648 DKVAKMYSDLRKESMATGSIPITVRHIESMIRMA--EAHARIHLRDYVIEDDVNMAIRVM 705

Qy        790 LHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFGAQQDTI 846
              |    : |:    ||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 LESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFGAQQDTI 765

Qy        847 EVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        766 EVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 811


RESULT 4
O42588
ID   O42588      PRELIMINARY;      PRT;   886 AA.
AC   O42588;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   MCM2p.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97357318; PubMed=9214647;
RA   Kubota Y., Mimura S., Nishimoto S., Masuda T., Nojima H., Takisawa H.;
RT   "Licensing of DNA replication by a multi-protein complex of MCM/P1
RT   proteins in Xenopus eggs.";
RL   EMBO J. 16:3320-3331(1997).
DR   EMBL; U44047; AAC60223.1; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   886 AA;  100260 MW;  3FEF85ADDE1CBC1D CRC64;

  Query Match           76.7%;  Score 3551.5;  DB 13;  Length 886;
  Best Local Similarity   78.5%;  Pred. No. 3.8e-247;
  Matches  706;  Conservative   57;  Mismatches   93;  Indels   43;  Gaps    8;

Qy         15 LTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRA 74
              : :||   ||| ||||||||||||||||||: |     |||||||||||| ||||||
Db         10 IATSPRAGSRR-DALTSSPGRDLPPFEDESEGMFGDGVVPEEEEDGEELIGDAMERDYRP 68

Qy         75 IPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTGSWPGLGACAVGSCM 127
              | ||| || |||  |:||||:||||:||||       |   |   |
Db         69 ISELDRYEVEGLD-DEEDVEDLTASQREAAEQSMRMRDREMGRELG-------------- 113

Qy        128 TAMRR-----------TRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVS 176
                |||           | |     |  |     ||||:||||||||||:|||:||:|||
Db        114 -RMRRGLLYDSDEEEEDRPARKRRMAERAAEGAPEEDEEMIESIENLEDMKGHTVRKWVS 172

Qy        177 MAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYF 236
              ||  ||||:|||||||||||| |||||||:|||||||:||| |||||||||||||||
Db        173 MAATRLEIYHRFKNFLRTHVDEHGHNVFKEKISDMCKENKESLPVNYEDLAAREHVLAYF 232

Qy        237 LPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIR 296
```

```
Db        233 LPEAPAEMLKIFDEAAKEVVLVMYPKYDRIAREIHVRISHLPLVEELRSLRQLHLNQLIR 292

Qy        297 TSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEE 356
Db        293 TSGVVTCCTGVLPQLSMVKYNCNKCNFILGPFFQSQNQEVRPGSCPECQSFGPFEINMEE 352

Qy        357 TIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLN 416
Db        353 TVYQNYQRITIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTGIYHNNYDGSLN 412

Qy        417 TANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYG 476
Db        413 TANGFPVFATVILANHITKKDDKVAVGELTDEDVKAIVALSKDERIGERIFASIAPSIYG 472

Qy        477 HEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTT 536
Db        473 HEDIKRGLALALFGGEAKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYVEKVASRAVFTT 532

Qy        537 GQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQ 596
Db        533 GQGASAVGLTAYVQRHPVTKEWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQ 592

Qy        597 SISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTV 656
Db        593 SISISKAGIVTSLQARCTVIAASNPIGGRYDPSLTFSENVDLTEPIVSRFDILCVVRDTV 652

Qy        657 DPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYA 716
Db        653 DPVQDEMLARFVVSSHIKHHPSSKD---IANGDAAEFALPNTFGVEALPQEVLKKYIMYA 709

Qy        717 KERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDR 776
Db        710 KEKIRPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA--EAHARMHLRDY 767

Qy        777 RRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLVAEQVT 833
Db        768 VVEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLAFRRDNNELLLFVLKQLIAEQVT 827

Qy        834 YQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892
Db        828 YQRNRYGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSDLFKMNKFTHDVKKKLIIQQF 886


RESULT 5
Q8JGR5
ID   Q8JGR5      PRELIMINARY;      PRT;   880 AA.
AC   Q8JGR5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   DNA replication licensing factor.
GN   MCM2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22035902; PubMed=12006978;
RA   Golling G., Amsterdam A., Sun Z., Antonelli M., Maldonado E., Chen W.,
RA   Burgess S., Haldi M., Artzt K., Farrington S., Lin S.-Y., Nissen R.M.,
RA   Hopkins N.;
RT   "Insertional mutagenesis in zebrafish rapidly identifies genes
RT   essential for early vertebrate development.";
RL   Nat. Genet. 31:135-140(2002).
DR   EMBL; AY099531; AAM28219.1; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.

SQ   SEQUENCE   880 AA;  98304 MW;  34C496C2ED14ABBC CRC64;

  Query Match             69.0%;  Score 3192.5;  DB 13;  Length 880;
  Best Local Similarity   73.4%;  Pred. No. 2.9e-221;
  Matches  658;  Conservative   64;  Mismatches  130;  Indels   45;  Gaps   16;

Qy         15 LTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRA 74
Db         10 MATSPTRGSRRGD-LTSSPGRDLPPFEDESEGLLGDTLPDEEDDDGEELIGDAMERDYRV 68

Qy         75 IPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCM------- 127
Db         69 VPELDRYEAEGLD-EDEDLSELSPSARAEAEAAMRR-RDREQGLGMGRIGRGLLYDSEDE 126

Qy        128 TAMRRT-RSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHH 186
Db        127 DDKRPTKRQRVLAERAAEGGAMEG-EDEEMIESIENL----GHE-----GPHGARVGFYG 176

Qy        187 R--------FKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLP 238
Db        177 RTTTGDLPPLKNFLRTHVDEHGHNVFKERISDMCKENKESLLVNYEDWLPRACV-GVFST 235

Qy        239 EAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTS 298
Db        236 KAPAEMLKIFDEAAKEVVLAMYPKYDRIAHEIHVRIGNLP----LGSLRQL---QLIPTS 288

Qy        299 GVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETI 358
Db        289 GVVTNCTGVLPQLGMVKYNCNKCNFILGPVFQSQNQEVKPGSCPECQSLGPFEINMEQTV 348

Qy        359 YQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTA 418
Db        349 YQNYQRITIQESPGKVAAGRLPRSKDAILLADLVDMCKPGDEIELTGIYHNNYDGSLNMA 408

Qy        419 NGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHE 478
Db        409 NGFPVFATVILANHIARKDEGVAVAELTDEDVKAIVALSKDERIGERIFASIGPFIYGHE 468

Qy        479 DIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQ 538
Db        469 DIKRGLALALFGGEAKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYVEKVASRAVFTTGQ 528

Qy        539 GASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSI 598
Db        529 GASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSI 588

Qy        599 SISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDP 658
Db        589 SISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDVLCVVRDTVDP 648

Qy        659 VQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKE 718
Db        649 VQDEMLARFVVGSHIKHHPSNK-EGGVA--GLEEVVLPNTFDVPPIPQELLRKYIIYAKE 705

Qy        719 RVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRR 778
Db        706 RVRPKLNQMDQDKVARIYSDLRKESMATGSIPITVRHIESMIRMA--EAHARMHLRDYVL 763

Qy        779 RQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQ 835
Db        764 EDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLAFRRDNNELLLFILKQLVSEQVSYQ 823

Qy        836 RNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQF 892
Db        824 RNRYGAQQDTIEIAEKDLVDKARQINIHSLSAFYDSDLFRSNKFSHDVKKKLIVQQF 880


RESULT 6
Q9XXI9
ID   Q9XXI9      PRELIMINARY;      PRT;   881 AA.
AC   Q9XXI9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Y17G7B.5 protein.
GN   Y17G7B.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smye R.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL023828; CAA19452.1; -.
DR   WormPep; Y17G7B.5; CE19038.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding.
SQ   SEQUENCE   881 AA;  99322 MW;  859E85B2838D5618 CRC64;

  Query Match             48.7%;  Score 2254;  DB 5;  Length 881;
  Best Local Similarity   52.4%;  Pred. No. 1.4e-153;
  Matches  465;  Conservative  129;  Mismatches  226;  Indels   68;  Gaps   15;

Qy         41 EDESEG---LLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT 97
Db         23 DDDVDGIDEMFNNEDEDPEDEEGENLFGDDMERDYREQPELDQYSESGMD-DASDVGSLS 81

Qy         98 ASRREAA------------------DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALP 138
Db         82 VSARRAAEREMAQRDQLLDDDALMYEDGDSEEVDTRRRGRG----------RRGR---- 126

Qy        139 ASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRT-HVD 197
Db        127 ----GDAADDDSVPMEEEDIPVDILENIRGRTIRDHVSDEAVAKEIERRFKNFLRSFHEP 182

Qy        198 SHGHNVFKERISDMCKENRESLVVNYEDLA--AREHVLAYFLPEAPAELLQIFDEAALEV 255
Db        183 GNKQTKYIQMIKSMAADNRESLEVSFTDLSDDNGEQNISYFLPEAPNEMLAIMDRAATEV 242

Qy        256 VLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVK 315
Db        243 VMNMYPFYSRVCNEIKVRISQLPVEEDIRMLRQVHLNMLIRTAGVVTIASGILPQLAVVK 302

Qy        316 YNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVA 375
Db        303 YDCVACGYLLGPFVQQNDEEVRPTICPSCQGKGPFELNVENTVYHNYQRITMQESPNKVA 362

Qy        376 ARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAK 435
Db        363 AGRLPRSKDVILLGDLCDSCKPGDEIEVTGVYTNNFDGSLNYKQGFPVFNTLIHANHITN 422

Qy        436 KDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKN 495
Db        423 KD-KMASDQLTDEDIKAIRELSQDPNISQRVFSSIAPSIYGHDDVKRAIALALFRGEAKN 481

Qy        496 PGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVS 555
Db        482 PGAKHRLRGDINVLLCGDPGTAKSQFLRYAAHIAPRSVLTTGQGASAVGLTAYVQRHPVT 541

Qy        556 REWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTV 615
Db        542 REWTLEAGAMVLADKGVCLIDEFDKMSDQDRTSIHEAMEQQSISISKAGIVTSLHARCTV 601

Qy        616 IAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRH 675
Db        602 IAASNPIGGRYNPTRTFAENVDLTEPILSRFDVLCVIRDSVDSVEDERLAKFVVGNHRTH 661

Qy        676 HPSNKE--EEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVA 733
Db        662 HPDAKKIVKEG---DELEEDQMDERTGVRLIPQDLLRKYIIYAREKCHPTLPEQHSEKFS 718

Qy        734 KMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE---- 789
Db        719 NIFAQMRKESMATGSVAITVRHVESMIRLS--EAHAKLHL-----RSYVNDEDCAAAIRV 771

Qy        790 -LHRHTEVQR---HRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTY--QRNRFGAQQ 843
Db        772 MLESFVNTQKASIMRMMKKTFSRHLTENRSANELLLFILKQLIRQQMHYATARAAAGTIL 831

Qy        844 DTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMILQQ 891
Db        832 QSVTIPESEFIEKAQQLRIENVKPFYTSEIFASNNFLYDPSKKTIVQE 879


RESULT 7
Q9BI22
ID   Q9BI22      PRELIMINARY;      PRT;   871 AA.
AC   Q9BI22;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   MCM2 protein (Fragment).
GN   MCM2.
OS   Dugesia japonica (Planarian).
OC   Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Seriata; Tricladida;
OC   Paludicola; Dugesiidae; Dugesia.
OX   NCBI_TaxID=6161;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Salvetti A., Rossi L., Deri P., Batistoni R.;
RT   "A MCM2-related gene is expressed in proliferating cells of intact and
RT   regenerating planarians.";
RL   Dev. Dyn. 0:0-0(0).
DR   EMBL; Y19170; CAC36296.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding.
FT   NON_TER       1      1
SQ   SEQUENCE   871 AA;  98504 MW;  80110EF8E7088461 CRC64;

  Query Match             46.0%;  Score 2132;  DB 5;  Length 871;
  Best Local Similarity   51.0%;  Pred. No. 9e-145;
  Matches  445;  Conservative  137;  Mismatches  217;  Indels   74;  Gaps   14;

Qy         36 DLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEE 95
Db         21 DDSPFEDETEQLLGAEYQDEESEE-EELINDSMERDYRRIPELDSYEAAGID-DESDVDE 78

Qy         96 LTASRREAADG-----------PCGTV-TGSWPGLGACAVGSCMTAMRRTRSALPASAAS 143
Db         79 ISQNERNVVEQELDRRDRLEAVASGDLRRGVIPDIFEDESSDAEASFRRQRMADKIASYD 138

Qy        144 GAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNV 203
Db        139 SAGETLG--SQQTYEAVENLEDTKGMSVAMWIQQPATKLEIRNRFFSFLRTFLNDEGKAV 196

Qy        204 FKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKY 263
```

```
Db        197 YSERIGTMAQANKKSLTINYEHLVAVEQALAYFLPEAPQGMLEIFDEVAFAVTLSRYPNY 256

Qy        264 DRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVL---PQLSMVKYNCNK 320
              ||  |||||||||||:|:|||||:|:||||||  :   |  ||   |||::|::|| :
Db        257 KTITESIHVRISHLPLVDEIRSLRQVHINQLIRTPEL---CQVVLLFCPQLNLVQFNCMR 313

Qy        321 CNFVLGPFCQSQN-QEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRL 379
              |   :||| |: :  |:|| :||:||| ||||:|:|:|:|:|||||  ||||| | | ||
Db        314 CGCTIGPFTQTDSTNEIKPSTCPDCQSNGPFEINVEKTLYKNYQRITGQESPGTVPAGRL 373

Qy        380 PRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK 439
              |||||||| ||||||  ||||::|||   || :||    ||||:| || |:| | |
Db        374 PRSKDAILLDDLVDSCKPGDEIDITGIYFIYYDRALNNKQCFPVFSTNILVNYVLKTDEH 433

Qy        440 VAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGK 499
              : :  :||||:  | :|:||::: ::|  ||||||||||:||  ||:||||  |  | |
Db        434 LILSGVTDEDIVNIQNLAKDERLFDRILRSIAPSIYGHENIKRAIALSLFGGVAKTKGQK 493

Qy        500 HKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWT 559
              : |||||||:|||||||||||||::|::: | :|||||||||| :|||| |:| |:|||
Db        494 LRGRGDINVLICGDPGTAKSQFLKFVEQLAPRCVFTTGQGASAVGLTAYVSRNPTSKEWT 553

Qy        560 LEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAA 619
              |||||||||:|||||||||||| ||||||||||||:|||||||||||||| |:||||
Db        554 LEAGALVLADKGVCLIDEFDKMNGQDRTSIHEAMEQQTISISKAGIVTSLQARSTIIAAA 613

Qy        620 NPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHP-- 677
              ||| |||| |  ||:|||| ||:|||||||||| |||:|| |||::|:|||:|||
Db        614 NPIAGRYDTSRNFSDNVDLTSPILSRFDILCVVRDIVDPIQDSMLAKYVIGSHMRHHGRL 673

Qy        678 SNKEEEGLANGSAAEPAMPNTY-----GVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKV 732
              | :|:: :     |  || ||       :: :| |:|:||||||::|| |:|:: :: |:
Db        674 SEEEKKEVVKRLEALGAMVNTSDSEAGDLQAIPLELLQKYIIYARQRVSPRLSKTNEGKL 733

Qy        733 AKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGELHR 792
              ::| :||| |   | | :|||: |||                  |    |     :||
Db        734 CEVYVELRKASKENGGIDVTVRYFESMI----------------RLSEAH----AKLHF 772

Qy        793 HTEVQRH---------------------RSMRKTFARYLSFRRDNNELLLFILKQLVAE 830
                  |                      ::   |  ||:::::::||::| |||:  :
Db        773 REVVNEEDVNMAIRVLLESFISTEKYSLMKNLENKFRHYLNYKKNDHELIMFTLKQMFHD 832

Qy        831 QVTYQRNRFGAQQDT-IEVPEKDLVDKARQINI 862
              |:|  |    :  :: ||: || |:|| |::||
Db        833 QLTTLRAEGKSLNNSIIEITEKSLLDKVRRMNI 865


RESULT 8
Q9LPD9
ID   Q9LPD9      PRELIMINARY;      PRT;   936 AA.
AC   Q9LPD9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   T12C22.19 protein.
GN   T12C22.19.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Liu S., Vaysberg M., Sakano H., Lee J., Lenz C., Pham P., Toriumi M.,
RA   Yu G., Chin C., Chiou J., Choi E., Chung M., Gonzalez A., Howng B.,
RA   Liu A., Altafi H., Brooks S., Buehler E., Chao Q., Conn L., Conway A.,
RA   Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M.,
RA   Nguyen M., Palm C., Shinn P., Southwick A., Davis R., Ecker J.,
RA   Federspiel N., Theologis A.;
RT   "The sequence of BAC T12C22 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC020576; AAF78275.1; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   936 AA;  105583 MW;  EDED5161964DBC69 CRC64;

  Query Match             42.0%;  Score 1945.5;  DB 10;  Length 936;
  Best Local Similarity   45.3%;  Pred. No. 2.9e-131;
  Matches  430;  Conservative  150;  Mismatches  268;  Indels  101;  Gaps   20;

Qy         12 NDPLT-SSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGT--------EGPLEEEEDGEE 62
              |:| : :|| :  || |  |  ::   | || |  : |        |   ||||:||:
Db          8 NEPSSPASPSSAGFNTDQLPISTSQNSENFSDEEEAAVDTQVIRDEPDEAEDEEEEGED 67

Qy         63 LIGDGMERDYRAIPELDAYEAEGL--ALDDE-DVEELTASRREAADGPC---------- 108
              |  |    ||| : | | ||: |:  ::||| |: :    || |||
Db         68 LFNDTFMNDYRKMDENDQYESNGIDDSVDDERDLGQAMLDRR-AADADLDARENRLANRK 126

Qy        109 ------GTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDG---------EED 153
                       :  |           :|:|: :|        |  ||          : |
Db        127 LPHLLHDNDSDDWN----------YRPSKRSRTTVP--PRGNGGDPDGNPPSSPGVSQPD 174

Qy        154 EQMIESIENLED-------------LKGHSVREWVSMAGPRLEIHHRFKNFLRTHV--- 196
              | :  :: :|              ::| ::||||     |  |  :||:|| |:|
Db        175 ISMTDQTDDYQDEDDNDDEAEFEMYRIQG-TLREWVMRDEVRRFIAKKFKDFLLTYVKPK 233

Qy        197 DSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVV 256
              : :|   :   |::|   |: || ::|::       :| :| :||  :|:: :| : :|:
Db        234 NENGDIEYVRLINEMVSANKCSLEIDYKEFIHVHPNIAIWLADAPQPVLEVMEEVSEKVI 293

Qy        257 LAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKY 316
              ::| |  |   |:||:::||: :::|::||:||| :||  ||||  :|| |||  |||
Db        294 FDLHPNYKNIHTKIYVRVTNLPVNDQIRNIRQIHLNTMIRIGGVVTRRSGVFPQLQQVKY 353

Qy        317 NCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAA 376
              :|||  ||||| |:   ||| ||| |||| ||| ||:|:|||:|||::  |||||| | |
Db        354 DCNKCGAVLGPFFQNSYSEVKVGSCSECQSKGPFTVNVEQTIYRNYQKLTIQESPGTVPA 413

Qy        377 RRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKK 436
              |||| |: ||| ||:|    |:|||:|||| ||:| |||| ||||||||: ||:| ||
Db        414 GRLPRHKEVILLNDLIDCARPGEEIEVTGIYTNNFDLSLNTKNGFPVFATVVEANYVTKK 473

Qy        437 DNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNP 496
              :  :  :|| ||   |  |||| :| |:|  ||||||||||||   |||:|||: ||
Db        474 QDLFSAYKLTQEDKTQIEELSKDPRIVERIIKSIAPSIYGHEDIKTALALAMFGGQEKNI 533

Qy        497 GGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSR 556
              |||::|||||||| ||||||||||||:||   ||::|||:||||| :|| | : ||:|
Db        534 KGKHRLRGDINVLLLGDPGTAKSQFLKYVEKTGQRAVYTTGKGASAVGLTAAVHKDPVTR 593

Qy        557 EWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVI 616
              ||||| |||||||||:|||||||||||||| ||||||||||||||||||||||||:||
Db        594 EWTLEGGALVLADRGICLIDEFDKMNDQDRVSIHEAMEQQSISISKAGIVTSLQARCSVI 653

Qy        617 AAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHH 676
              |||||:|||| | :|::||:||:||:|||||||:| |||| ||||| ||| || :
Db        654 AAANPVGGRYDSSKSFAQNVELTDPILSRFDILCVVKDVVDPVTDEMLAEFVVNSHFKSQ 713

Qy        677 P------SNKEEEGLANGS-AAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQ 729
              |       .:  |:|:   | : :|        | ||| :|||: |:|  | ||| ::|
Db        714 PKGGKMEDSDPEDGIQGSSGSTDP--------EVLPQNLLKKYLTYSKLYVFPKLGELDA 765

Qy        730 DKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE 789
              |:  :|::||:|||    : |  ||:|||       ||    |     ||:    |
Db        766 KKLETVYANLRRESMNGQGVSIATRHLESMIRMS--EAHARMHL-----RQYVTEEDVNM 818

Qy        790 -----LHRHTEVQR---HRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFGA 841
                  |       |:    |::|::| ||:::::| | ||| :||:||   : ::    |:
Db        819 AIRVLLDSFISTQKFGVQRTLRESFKRYITYKKDFNSLLLVLLKELVKNALKFEEIISGS 878

Qy        842 QQ--DTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMI 888
```

```
Db        879 NSGLPTIEVKIEELQTKAKEYDIADLRPFFSSTDFSKAHFELDHGRGMI 927

RESULT 9
Q8H0G9
ID   Q8H0G9      PRELIMINARY;      PRT;   865 AA.
AC   Q8H0G9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   MCM protein-like protein.
GN   B37.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Samsun;
RA   Kyo M.;
RT   "Cloning of cDNAs associated with embryogenic dedifferentiation of
RT   immature pollen grains in Nicotiana tabacum.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032541; BAC53939.1; -.
SQ   SEQUENCE   865 AA;  98190 MW;  1779169B1B96C5E9 CRC64;

  Query Match             40.8%;  Score 1887.5;  DB 10;  Length 865;
  Best Local Similarity   47.1%;  Pred. No. 3.8e-127;
  Matches  414;  Conservative  150;  Mismatches  258;  Indels   57;  Gaps   17;

Qy         42 DESEGLLGTEGPLEEEEDGEELIGD----GMERDYRAIPELDAYEAEGLALDDEDVEELT 97
Db          5 DQYES-LGLDDSMEDERDLDQIMADRRAAEVELDTRDVQVTNRKLPQLLHDQDTDDDNYR 63

Qy         98 ASRREAAD-GPCGTV-----TGSWPGLGACAVGSCMTAMRRTRSA--LPASAASGAGTED 149
Db         64 PSKRTRADFRPTNTQRNFDDTDAMP--------SSPGASQRVNSSQDVPMTDQT---DDD 112

Qy        150 GEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVD---SHGHNVFKE 206
Db        113 AYEDDENDEGEFEMYRVQG-TLREWVTRDEVRRFIAKKFKEFLLTYVNPKSEHGDFEYLR 171

Qy        207 RISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRI 266
Db        172 QINEMVSVNKCSLEIDYKQFIYVHPNIAIWLADAPQSVLEVMEEIANKVVFDLHPNYKQI 231

Qy        267 TNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLG 326
Db        232 HQKVYVRITNLPVYDQIRNIRQIHLNTMIRIGGVVTRRSGVFPQLQQVKYDCNKCGAILG 291

Qy        327 PFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAI 386
Db        292 PFFQNSYSEVKVGSCPECQSKGPFTVNVEQTIYRNYQKLTLQESPGIVPAGRLPRYKEVI 351

Qy        387 LLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELT 446
Db        352 LLNDLIDCARPGEEIEVTGVYTNNFDLSLNTKNGFPVFATVIEANYVTKKQDLFSAYKLT 411

Qy        447 DEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDI 506
Db        412 QEDKEEIEKLAKDPRIGERISKSIAPSIYGHEDIKTALALAMFGGQEKNVEGKHRLRGDI 471

Qy        507 NVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALV 566
Db        472 NILLLGDPGTAKSQFLKYVEKTGQRAVYTTGKGASAVGLTAAVHKDPVTREWTLEGGALV 531

Qy        567 LADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRY 626
Db        532 LADRGICLIDEFDKMNDQDRVSIHEAMEQQSISISKAGIVTSLQARCSVIAAANPIGGRY 591

Qy        627 DPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHP--SNKEEEG 684
Db        592 DSSKTLTQNVELTDPIISRFDVLCVVKDVVDPVIDEMLAKFVVDSHFRSQAKGATLDEKS 651

Qy        685 LANG-----SAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDL 739
Db        652 FTDSRDDARAAMAPTDP-----EIIPQELLKKYITYAKLNVFPKLHDGDLDKLTQVYAEL 706

Qy        740 RKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE-----LHRHT 794
Db        707 RRESSHGQGVPIAVRHIESMIRMS--EAHARMHL-----RQHVTQEDVDMAIRVLLDSFI 759

Qy        795 EVQR---HRSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFG--AQQDTIEVP 849
Db        760 STQKFGVQKALQKSFKKYMTYKKDFNAIILHLLRGLVNDAMQFEEIVSGSTANLDHIDIK 819

Qy        850 EKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRKMI 888
Db        820 VDELQSKALDYGITDLKAFFTSNDFSKANFELDKERGII 858

RESULT 10
O42723
ID   O42723      PRELIMINARY;      PRT;   889 AA.
AC   O42723;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   DNA replication licensing factor.
GN   NIMQ.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R153;
RX   MEDLINE=98070550; PubMed=9407133;
RA   Ye X.S., Fincher R.R., Tang A., McNeal K.K., Gygax S.E., Wexler A.N.,
RA   Ryan K.B., James S.W., Osmani S.A.;
RT   "Proteolysis and tyrosine phosphorylation of p34(cdc2)/Cyclin B. The
RT   role of mcm2 and initiation of dna replication to allow tyrosine
RT   phosphorylation of p34.";
RL   J. Biol. Chem. 272:33384-33393(1997).
DR   EMBL; AF014813; AAB94861.1; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 2.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   889 AA;  99968 MW;  62BDA1118DDB5281 CRC64;

  Query Match             40.3%;  Score 1864.5;  DB 3;  Length 889;
  Best Local Similarity   46.7%;  Pred. No. 1.8e-125;
  Matches  413;  Conservative  114;  Mismatches  249;  Indels  109;  Gaps   16;

Qy          2 ASSPAQRRRGNDPLTSSPGRSSRRTD------ALTSSPGRDLPPF----EDESEGLLGTE 51
Db          8 STSAANRGLGN----LNRRKRSREPDDDASSVVPPSSPPSSPPMLPFDEDEDEP--DEE 61

Qy         52 GPLE--------EEEDGEELIGDGMERDYRAIPELDAYEAEGLAL----------DD--- 90
Db         62 AELDIDDIEEMAEDEDGIDLFGDTFERDYRGGKD-DRYRHDTLTTMSKKKSTLVPDESWR 120

Qy         91 ----EDVEELTASRREAA---DGPCGTVTGSWPGLGACAVGSCMTAM-RRTRSALPASAA 142
Db        121 LAWIEEIENLLGRRMPAAFLQDDEDGDID-------------LTAQPRRRRHAYDEDRD 166

Qy        143 SGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHN 202
Db        167 DIEMADDGLEE----LSLEELVDIKSSNITDWVTQPQVLRSIYREFKAFLTEFTDPTGSS 222

Qy        203 VFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPK 262
```

```
                |:  :|  : : |  || |:|  |:: : || |||   | |:|::||: ||:| |  ||:
Db        223 VYGNKIKTLGEVNSASLEVSYTHLSSTKAVLGYFLANEPTEVLKVFDQVALDVTLFHYPQ 282

Qy        263 YDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCN 322
              |  | | |||||: ||:|  || ||| ||| |:| |||||  ||| |||  | : | |||
Db        283 YHDIHNEIHVRITDLPIVYTLRQLRQSHLNCLVRVSGVVTRRTGVFPQLKYVMFICQKCN 342

Qy        323 FVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRS 382
                |||: |  : |||   |  ||| ||| #: |:|:|:|||:: :||||| | | ||||
Db        343 MTLGPYQQESSSEVKVTMCQNCQSRGPFTVSSEKTVYRNYQKLTLQESPGSVPAGRLPRQ 402

Qy        383 KDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAV 442
              :: ||||||:||   |||||:|||| |:||  ||  |||||||:| |||| |  :: |
Db        403 REVILLADLIDSAKPGDEIEITGIYRNSYDAQLNNKNGFPVFATIIEANHVIKSHDQQAG 462

Qy        443 GELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKV 502
              :||:|| : | :||:|  | |||  |||||||||:|:|:|   ||:|||| |  ||  :
Db        463 FQLTEEDEREIRALSRDPDIVEKIIRSIAPSIYGHQDVKTAVALSLFGGVRKETQGKMSI 522

Qy        503 RGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEA 562
              |||||||| |||||||| |||:|| : ||:| |||||||| :|| |:| |:: |||||
Db        523 RGDINVLLLGDPGTAKSQILKYVEKTAHRAVFATGQGASAVGLTASVRRDPLTSEWTLEG 582

Qy        563 GALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPI 622
              |||||||| ||||||||||||||||||||||:|||||||||:||||| |:||||||
Db        583 GALVLADRGTCLIDEFDKMNDQDRTSIHEAMEQQTISISKAGIVTTLQARCAVVAAANPI 642

Qy        623 GGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSN--K 680
              ||||: |  ||||| |||||:|||||||||| ||| :|| || ||: || | :||   :
Db        643 GGRYNSSAPFSENVQLTEPILSRFDILCVVRDLVDPSEDERLASFVIESHHRANPSKPLR 702

Qy        681 EEEGLANGSAAEPAMPNTYGVE----------------------------PLPQEVLKKY 712
              :| |    :  |      | ::                           :|||:|:||
Db        703 DEHGNLINADGERIDEEGYRIDDDGIRLPLRPDEIEARNAARDNDDEKEGEIPQELLRKY 762

Qy        713 IIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSAG 772
              |:||:|   ||| |:||||||::::|:|:||:|||: ||||||:|::          :
Db        763 ILYAREHCRPKLYQIDQDKVARLFADMRRESLATGAYPITVRHLEAI-------MRIAEA 815

Qy        773 LCDRRRRQHGHPRDAGELHRHT--------EVQRHRSMRKTFARY 809
               |  |  ::   :|       |        :|   |:: : ||:|
Db        816 FCKMRLSEYCSAQDIDRAIAVTVDSFIGSQKVSAKRALSRAFAKY 860


RESULT 11
Q9BWF4
ID   Q9BWF4      PRELIMINARY;      PRT;   365 AA.
AC   Q9BWF4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC000300; AAH00300.1; -.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   365 AA;  41578 MW;  DCA4594EB30EB988 CRC64;

  Query Match             34.7%;  Score 1604.5;  DB 4;  Length 365;
  Best Local Similarity   89.1%;  Pred. No. 2.5e-107;
  Matches  326;  Conservative    5;  Mismatches   30;  Indels    5;  Gaps    2;

Qy        530 SRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSI 589
              :|||||||||||| :|||||||||||||||||||||||||||||||||||||||||||||
Db          2 TRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSI 61

Qy        590 HEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDIL 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 HEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDIL 121

Qy        650 CVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVL 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVL 181

Qy        710 KKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHP 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 KKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA--EAHA 239

Qy        770 SAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLFILKQ 826
                | |       :      |    : |:    ||||||||||||||||||||||||||
Db        240 RIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLFILKQ 299

Qy        827 LVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRK 886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 LVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHDLKRK 359

Qy        887 MILQQF 892
              ||||||
Db        360 MILQQF 365


RESULT 12
Q8SS42
ID   Q8SS42      PRELIMINARY;      PRT;   780 AA.
AC   Q8SS42;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   DNA replication licensing factor MCM2.
GN   ECU04_0850.
OS   Encephalitozoon cuniculi.
OC   Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX   NCBI_TaxID=6035;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GB-M1;
RA   Genoscope;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GB-M1;
RX   MEDLINE=21576510; PubMed=11719806;
RA   Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA   Prensier G., Barbe V., Peyretaillade E., Brottier P., Wincker P.,
RA   Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA   Weissenbach J., Vivares C.P.;
RT   "Genome sequence and gene compaction of the eukaryote parasite
RT   Encephalitozoon cuniculi.";
RL   Nature 414:450-453(2001).
DR   EMBL; AL590444; CAD25272.1; -.
DR   InterPro; IPR001220; Lectin_legB.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
SQ   SEQUENCE   780 AA;  88251 MW;  E517A434D4716033 CRC64;

  Query Match             32.9%;  Score 1525;  DB 5;  Length 780;
  Best Local Similarity   39.8%;  Pred. No. 4.5e-101;
```

```
  Matches  351;  Conservative  149;  Mismatches  238;  Indels  144;  Gaps   25;

Qy         41 EDES----EGLLGTEGPLE-EEEDGEELIGDGMERDYRAIPELDAYEAEGL-----ALDD 90
              ||||    | :|   |  | | || :|:      ||       | ||:||:     | :
Db          8 EDESMSEEERILDAGGVQELESEDMDEI------RD-------DLYESEGMYEESEATEH 54

Qy         91 EDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDG 150
               |:|:| :                                 :  ::  |     | |  |
Db         55 SDIEDLIS------------------------------QEVKNVRPDERKGGNG--DA 80

Qy        151 EED-EQMIESIENLED------LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNV 203
              | | | :::: |  :|       :   ||:::           || |         |:
Db         81 EYDAEDILKASEVTKDFFEQDFFRAKVVRKFI-----------RFFN-------EFGNKK 122

Qy        204 FKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKY 263
              : :||  || || ||: |:| |:       |   | :     ::: | |  :||   :| |
Db        123 YAKRIRQMCTENLESIEVSYLDIEEESIDLLRLLNQHAEMTIEVMDRALSDVVRMHFPNY 182

Qy        264 DRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNF 323
              |   :| |:  ||: : :|||| || :|:| |||||  :|| |  |:||::| ||
Db        183 HMIKPKVHSRVVDLPVCDSVRSLRNRHLGKLVRVSGVVTRRSGVFPLYSIVKFSCLKCRS 242

Qy        324 VLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSK 383
              | |||  |     ||  | |||| ||| ||  ||:|:::|:: ||| || |    |||||
Db        243 VFGPFVAS---SFKPTHCFECQSKGPFTVNTSETVYKDFQKLTIQEIPGSVPPGSLPRSK 299

Qy        384 DAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVG 443
              : :|  ||:|    |:|:|:||:| ||:: |||  |||||| ||| |: | |:  |:
Db        300 EVLLFYDLIDCAKPGEEVEVTGVYKNNFNVSLNIRNGFPVFFTVIEASSVVKRAGKI--- 356

Qy        444 ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVR 503
              |:||:||: |  : :  :|   :  |||||:||| ::||  |||: ||  :     |::|
Db        357 EMTDDDVREIKKMGRHPEIKRIVINSIAPSVYGHAEVKRAIALAMLGGVARE-STSHRIR 415

Qy        504 GDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAG 563
              ||||||| |||| ||||||:|:|  | ||:  ||||||:| :|| |:: || :|||| |
Db        416 GDINVLLLGDPGMAKSQFLRYVESTSHRAVLATGQGASSVGLTASVRKDPVVKEWTLEGG 475

Qy        564 ALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIG 623
              ||||||:|:|||||||||: |||||||||||||||||||| :| |||:|||||||:
Db        476 ALVLADKGICLIDEFDKMNEHDRTSIHEAMEQQSISISKAGIVATLHARCSVIAAANPMR 535

Qy        624 GRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEE 683
              |||: ||||::||:|::|||||||||||||:| :|  :||  |:||: ||
Db        536 GRYNGSLTFAQNVNLSDPIISRFDILCVVKDAIDAGEDEKTAKFVIESH---------- 584

Qy        684 GLANGSAAEPAMPNTYGVEP----LPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDL 739
                  |   :|      | :|    :  |:|:|||:||:  | |  | :| :|:: :| :|
Db        585 ---EGGEEKPD-----GFDPKRMMMGHELLRKYILYARTNVVPAFNDVDMEKISSLYLEL 636

Qy        740 RKESMATGSIPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGE-----LHRHT 794
              ||||: :| :|:||||:||:          |      |  :     | |      |
Db        637 RKESLPSG-LPVTVRHVESI-------VRISEAFAKMRLSRVVSVEDIDEAISVVLDSFM 688

Qy        795 EVQRH---RSMRKTFARYLSFRRDNNELLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEK 851
              |::   :|:|| | :|  | ::| ::|:|:||::  |::   |::      :: | |
Db        689 GAQKYSMSKSLRKKFVKY--FNKNNIDVLVFLLKEMFNEKMKAFRSQ------SVSVDEF 740

Qy        852 DLVDKARQINIHNLSA---FYDSELFRMNKFSHDLKRKMILQ 890
              :     |:|:    |    ||   ||: : |  | : ::||:
Db        741 E-----RRISSFGFSIPSNFYSCGLFKDSGFRLDREARLILR 777


RESULT 13
Q9U446
ID   Q9U446      PRELIMINARY;      PRT;   883 AA.
AC   Q9U446;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Minichromosome maintenance protein 2 homolog.
GN   MCM2.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM1:IMSS;
RA   Lohia A.;
RT   "Characterization of Mcm proteins of Entamoeba histolytica.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF203971; AAF17244.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS00847; MCM_1; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding.
SQ   SEQUENCE   883 AA;  100124 MW;  8C2C2574DD1E7D71 CRC64;

  Query Match             27.6%;  Score 1277;  DB 5;  Length 883;
  Best Local Similarity   39.6%;  Pred. No. 4.3e-83;
  Matches  303;  Conservative  124;  Mismatches  263;  Indels   76;  Gaps   15;

Qy         41 EDESEGLLGT-------EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDV 93
              :|| | |           |   | | |   || : :| :   :      | |     ||::
Db         94 DDEGEDLFNPKLMDEDYERSEESEYDDPNLINNEIEHNEGVV------EIEARRKADEEI 147

Qy         94 EELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTED---- 149
              :   |  |:  |    ::                :    | || :          :|
Db        148 DRRNALMRKVIDEEDSSIE---------EFNLLVEQQHRRRSEIEKLTMK----QDTDIV 194

Qy        150 ----GEEDEQMIESIE--NLEDLKGHSVREWVSMAGPRLEIHHR---------------- 187
                  |  :|:: |  :  :| :     :  :| |  :|::: :
Db        195 IPQFGLNEEEVKERFKQREIEKMPIEQFQMMMSEAESKLQMYTKEHIKDAIFLEDVHDLL 254

Qy        188 ---FKNFL--RTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPA 242
                 | ||   |  ::  : |:   |  |   |  || |::  ||    :|| ::   |:
Db        255 VETLKIFLLYYTETETSKNVVYIISIRKMVSRNSGSLDVSFLHLAKSSKLLAQWIVLCPS 314

Qy        243 ELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVT 302
               :: || ||| :  | :|| |  |:  ::|||         || ||  |:| |||  |:||
Db        315 SVIPIFSEAATQATLLLYPDYKDISTQVNVRIVDYTTRIALRDLRHCHINTLIRVVGIVT 374

Qy        303 SCTGVLPQLSMVKYNCNKCNFVLGPFC--QSQNQEVKPGSCPECQSAGPFEVNMEETIYQ 360
                 | : |||  ||| |: |   |||:   :  |:  :    |  ||| ||| :::: ||||
Db        375 RVTAIFPQLKAVKYICSVCQARLGPYFINKEMNKVPQLQVCTVCQSKGPFSIDVQNTIYQ 434

Qy        361 NYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANG 420
              |||:| ||| |  |:|  :||:|| ||| ||:|    |:||:: |    ||:  ||   |
Db        435 NYQKITIQEPPNSVSAGNVPRTKDVILLGDLIDKAQPGEEIDINGNVCPNYETGLNRNFG 494

Qy        421 FPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDI 480
              |||| |||  | : |:   |    :| |: : |  |: : || : |  ||||:||||:
Db        495 FPVFCTVIEVNTIEKRSGDVISTTITHEEEQEIRRLANNPQIFQIIINSIAPAIYGHDAS 554

Qy        481 KRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQ 538
              |  |||||||| :    | |: |||||||| |||||||| ||| :|:: ||:||||:
Db        555 KAAIALALFGGEQRVLVDKGNHRTRGDINVLLLGDPGTAKSQLLKYSQKLAPRAVFTTGR 614

Qy        539 GASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSI 598
              |::|| :|| |::  :: || || ||||||| ||||||||||:|||||||||||||||
Db        615 GSTAVGLTAAVKKDSMNGEWALEGGALVLADEGVCLIDEFDKMDDQDRTSIHEAMEQQSI 674

Qy        599 SISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDP 658
              ||||||||||:|||:|||||||  |:|:|:   ::||:|||||||||:: :||| ||
Db        675 SISKAGIVTSLKARCSVIAAANPKTGKYNPNKNLNQNVNLTEPIISRFDLIMIVRDVVDY 734

Qy        659 VQDEMLARFVVGSHVRHHP-SNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAK 717
              |:   ||:||| ||  :|| :::: | :|        |      :   :|||| ||:
Db        735 VKVYKLAQFVVESHSINHPEASQKRESIA---------PIVNKTNIISHVLLKKYIAYAR 785

Qy        718 ERVHPKLN-QMDQDKVAKMYSDLRK--ESMATGSIPITVRHIESMS 760
```

```
Db        786 QNCHPKWSGTVGSQMIQQAYIEMRKCCDKYHTGQ--VTARQIEAIN 829


RESULT 14
Q8ILR7
ID   Q8ILR7      PRELIMINARY;      PRT;   971 AA.
AC   Q8ILR7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   DNA replication licensing factor MCM2.
GN   PF14_0177.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014818; AAN36789.1; -.
SQ   SEQUENCE   971 AA;  111496 MW;  E5670CA02AA1C208 CRC64;

  Query Match             27.4%;  Score 1267;  DB 5;  Length 971;
  Best Local Similarity   34.9%;  Pred. No. 2.6e-82;
  Matches  336;  Conservative  147;  Mismatches  317;  Indels  162;  Gaps   28;

Qy         40 FEDESEGLLGTEGPLEEEEDGEELIGDGM-ERDYRAIPELDAYEAE-----GLALDDEDV 93
              :: : | ||  || | |||   || |: : | |   |   | : |       | ||::|
Db         16 YDIDEEDLLEDEGRLNEEERQAELEGESLFEEDEGFIFGADDEKKEMQKLRNLGLDNDDY 75

Qy         94 E----------ELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRR-----TRSALP 138
              :          |     | ||:                      |   |:      ::
Db         76 DDDFIDDELDYEDNLKARRAAE------------------RHMQMQRKQEGKYEKNKFW 116

Qy        139 ASAASGAGTEDGEED------EQMIESIENLE---------DLKG-HSVREWVSMAGPRL 182
              :       :| |||      |:: :  |||          ||    | :  :|:    :
Db        117 KTLEDQLEGDDEEEDIFDKVAEKVAKRRENLHLTAEETDIPDLSNLESAKICLSVNPKDV 176

Qy        183 EIHHR--------FKNFL-----------------RTHVDSH---GHNVFKERISDMCK 213
                 |         |: ||                | : |    |  : ::|  | .
Db        177 VFDERYQQAADTCFRYFLHKFSLKDSMGLNIDESNTELHEEEHMNSSHQYYIDKIEKMIL 236

Qy        214 ENRESLVVNYEDLAAREHV--LAYFLPEAPAELLQIFDEAALEVVLAMYPKY--DRITNH 269
              :: :|:|: : |  : |   |  ::   | ::|::  |  :     : ||    ||
Db        237 NDKHTLIVSAKHL-IQFHCENLVQWIEFKPEQILEVLHECLMVEAYRISPKLYKGRICK- 294

Qy        270 IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGP-- 327
               | :   |   :|:|:||   || ||: :||      |||:| ::   || |:  |
Db        295 --VVLRDWPYSTQLRNLRCTELNTLIKVTGVCIKRGYVLPKLRVMYLKCNSCDTTLSEVP 352

Qy        328 -FCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAI 386
               :     : | |  || |||| | |:  :| | :||:| :|||| | | | || :: :
Db        353 IYFADGKKPVLPRRCPHCQSA-TFSVDRIKTAYTDYQKITLQESPCSVPAGRAPRQREVV 411

Qy        387 LLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELT 446
              :  ||||    |:|:|: |||   ||  ||   |||:  | | ||:: :|:: : : |||
Db        412 VTGDLVDKVKPGEEVEVLGIYKTKYDIGLNIKYGFPILQTEIEANNIERKED-IQLSELT 470

Qy        447 DEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPK------------ 494
              ::|:| |  |||| | |:|  ||||:|:||:|||   | |||||  |

Db        471 EDDIKDILKLSKDPNIRERIITSIAPAIWGHKDIKTSIAYALFGGVQKGGDKSFSKNNET 530

Qy        495 ------------NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASA 542
                         |  | | :|||||||| |||  ||| |:|: | : | ::|||:||||
Db        531 NNFGVQNRDILNNFKGGHTIRGDINVLLLGDPGLGKSQVLQYVHKTNLRTVYTTGKGASA 590

Qy        543 VAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISK 602
              | :|| |::   : ||||| ||||||| |:|:|||||| |:|| |||||||||||||||
Db        591 VGLTAGVRKDHTTNEWTLEGGALVLADEGICIIDEFDKMTDKDRVSIHEAMEQQSISISK 650

Qy        603 AGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDE 662
              |||||:|:||| |||||||| |||||||| |||||::||:||||:: |:||  :  :|
Db        651 AGIVTTLRARCAVIAAANPIYGRYDPSLTFKENVDLSDPILSRFDLITVLRDIPNVDEDF 710

Qy        663 MLARFVVGSHVRHHPS-------NKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIY 715
               || :|| :|   ||         |  | | |       : ::   ||:||::|:|||||
Db        711 YLAEYVVTNHQLSHPKLENTQNYQKRIENLKN------VIVSSSAYEPIPQDLLQKYIIY 764

Qy        716 AKERVHPKLNQMD----QDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSA 771
              |:    | |: :       |::  || :|::: |:|  |:|:|||||:       |
Db        765 ARTNCKPSLSDVPYAEISAKLSNFYSRVRQKASASGGYPLTLRHIESIIRIAEANA---- 820

Qy        772 GLCDRRRRQHG-HPRD-----AGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
                 : |  |  : :|     |  |  :   ||    : : | |||| :  |   |:|
Db        821 ----KMRLSHQIYSKDVDYAIATLLESYVSCQRFAVAKQLSKEFARYRALFRGGYEVLRE 876

Qy        823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELF----RMNK 878
              :|:: | : :  |  :  : :|     |     :|  :| :|:  |    :|      ||
Db        877 LLRRTVQQMIQRQNLKNASAKDFQNDSESGTESEAEILNPNNV--FLPLHIFMKTAMQNK 934

Qy        879 FS 880
              ||
Db        935 FS 936


RESULT 15
Q9GR05
ID   Q9GR05      PRELIMINARY;      PRT;   971 AA.
AC   Q9GR05;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   DNA replication licensing factor MCM2.
GN   MCM2.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7A;
RA   Li J.-L.;
RT   "Identification of a novel gene encoding DNA replication licensing
RT   factor MCM2 homolog from the human malaria parasite Plasmodium
RT   falciparum.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF095948; AAG37990.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001208; MCM.
DR   Pfam; PF00493; MCM; 1.
DR   ProDom; PD001041; MCM; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00350; MCM; 1.
DR   PROSITE; PS50051; MCM_2; 1.
KW   ATP-binding.
SQ   SEQUENCE   971 AA;  111525 MW;  DC1F08EF395FA92E CRC64;

  Query Match             27.3%;  Score 1262;  DB 5;  Length 971;
  Best Local Similarity   34.8%;  Pred. No. 6e-82;
  Matches  335;  Conservative  148;  Mismatches  317;  Indels  162;  Gaps   28;

Qy         40 FEDESEGLLGTEGPLEEEEDGEELIGDGM-ERDYRAIPELDAYEAE-----GLALDDEDV 93
              :: : | ||  || | |||   || |: : | |   |   | : |       | ||::|
Db         16 YDIDEEDLLEDEGRLNEEERQAELEGESLFEEDEGFIFGADDEKKEMQKLRNLGLDNDDY 75
```

```
Qy          94 E----------ELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRR-----TRSALP 138
               :          |     | ||:                     |   |:      ::
Db          76 DDDFIDDELDYEDNLKARRAAE------------------RHMQMQRKQEGKYEKNKFW 116

Qy         139 ASAASGAGTEDGEED------EQMIESIENLE---------DLKG-HSVREWVSMAGPRL 182
                :       :| |||      |:: :  |||          ||    | :  :|:     :
Db         117 KTLEDQLEGDDEEEDIFDKVAEKVAKRRENLHLTAEETDIPDLSNLESAKICLSVNPKDV 176

Qy         183 EIHHR--------FKNFL-----------------RTHVDSH---GHNVFKERISDMCK 213
                    |       |: ||                     | : |    |  : ::|  |
Db         177 VFDERYQQAADTCFRYFLHKFSLKDSMGLNIDESNTELHEEEHMNSSHQYYIDKIEKMIL 236

Qy         214 ENRESLVVNYEDLAAREHV--LAYFLPEAPAELLQIFDEAALEVVLAMYPKY--DRITNH 269
                :: :|:|: : |  : |   |  ::   | ::|::  |  :     : ||    ||
Db         237 NDKHTLIVSAKHL-IQFHCENLVQWIEFKPEQILEVLHECLMVEAYRISPKLYKGRICK- 294

Qy         270 IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGP-- 327
               | :   |   :||:||   || ||: :||      |||:| ::   || |:  |
Db         295 --VVLRDWPYSTQLRNLRCTELNTLIKVTGVCIKRGYVLPKLRVMYLKCNSCDTTLSEVP 352

Qy         328 -FCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAI 386
                :     : | |  || ||||  | |:  :| | :||:| :||||  | | | || :: :
Db         353 IYFADGKKPVLPRRCPHCQSA-TFSVDRIKTAYTDYQKITLQESPCSVPAGRAPRQREVV 411

Qy         387 LLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELT 446
               :  ||||    |:|:|: |||   ||  ||   |||:  | | ||:: :|:: : : |||
Db         412 VTGDLVDKVKPGEEVEVLGIYKTKYDIGLNIKYGFPILQTEIEANNIERKED-IQLSELT 470

Qy         447 DEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPK------------ 494
               ::|:| |  ||||  | |:|  ||||:|:||:|||   | |||||  |
Db         471 EDDIKDILKLSKDPNIRERIITSIAPAIWGHKDIKTSIAYALFGGVQKGGDKSFSKNNET 530

Qy         495 ------------NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASA 542
                          |. | | :|||||||| |||| |||  |:|: | : | ::|||:||||
Db         531 NNFGVQNRDILNNFKGGHTIRGDINVLLLGDPGLGKSQVLQYVHKTNLRTVYTTGKGASA 590

Qy         543 VAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISK 602
               | :|| |::   : ||||| ||||||| |:|:|||||||  |:|| |||||||||||||||
Db         591 VGLTAGVRKDHTTNEWTLEGGALVLADEGICIIDEFDKMTDKDRVSIHEAMEQQSISISK 650

Qy         603 AGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDE 662
               |||||:|:||| |||||||| |||:|||| |||||::||:||||:: |:||  :  :|
Db         651 AGIVTTLRARCAVIAAANPIYGRYNPSLTFKENVDLSDPILSRFDLITVLRDIPNVDEDF 710

Qy         663 MLARFVVGSHVRHHPS-------NKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIY 715
                || :|| :|   ||          |  | | |       : ::   ||:||::|:|||||
Db         711 YLAEYVVTNHQLSHPKLENTQNYQKRIENLKN------VIVSSSAYEPIPQDLLQKYIIY 764

Qy         716 AKERVHPKLNQMD----QDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRAHPSA 771
               |:    | |: :       |::  || :|::: |:|  |:|:|||||:        |
Db         765 ARTNCKPSLSDVPYAEISAKLSNFYSRVRQKACASGGYPLTLRHIESIIRIAEANA---- 820

Qy         772 GLCDRRRRQHG-HPRD-----AGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
                : |  |  : :|      |  |  :   ||    : : | |||| :  |   |:|
Db         821 ----KMRLSHQIYSKDVDYAIATLLESYVSCQRFAVAKQLSKEFARYRALFRGGYEVLRE 876

Qy         823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELF----RMNK 878
               :|:: | : :  |  :  : :|     |     :|  :| :|:  |    :|      ||
Db         877 LLRRTVQQMIQRQNLKNASAKDFQNDSESGTESEAEILNPNNV--FLPLHIFMKTAMQNK 934

Qy         879 FS 880
               ||
Db         935 FS 936
```

Search completed: September 12, 2003, 13:13:50
Job time : 59 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 13:12:35 ; Search time 20 Seconds
(without alignments)
1887.064 Million cell updates/sec

Title: CAA47749
Perfect score: 4630
Sequence: 1 MASSPAQRRRGNDPLTSSPG.....LFRMNKFSHDLKRKMILQQF 892

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	774	16.7	720	4	US-09-604-605-2	Sequence 2, Appli
2	140	3.0	918	4	US-09-198-452A-1072	Sequence 1072, Ap
3	128.5	2.8	1515	4	US-09-328-352-6100	Sequence 6100, Ap
4	126.5	2.7	871	4	US-09-134-001C-3979	Sequence 3979, Ap
5	125	2.7	15281	2	US-08-471-119A-2	Sequence 2, Appli
6	121.5	2.6	516	4	US-09-252-991A-24044	Sequence 24044, A
7	121	2.6	875	4	US-09-107-532A-4537	Sequence 4537, Ap
8	121	2.6	1162	4	US-09-134-001C-4008	Sequence 4008, Ap
9	118.5	2.6	1589	3	US-08-755-587-189	Sequence 189, App
10	117.5	2.5	1698	3	US-09-315-793-12	Sequence 12, Appl
11	117	2.5	616	2	US-09-001-826-5	Sequence 5, Appli
12	117	2.5	616	4	US-09-264-854-5	Sequence 5, Appli
13	116	2.5	771	4	US-09-252-991A-17243	Sequence 17243, A
14	116	2.5	1151	3	US-08-840-006-6	Sequence 6, Appli
15	116	2.5	1200	3	US-08-840-006-5	Sequence 5, Appli
16	115	2.5	780	4	US-09-252-991A-18846	Sequence 18846, A
17	115	2.5	1481	4	US-09-231-899-70	Sequence 70, Appl
18	113	2.4	1105	3	US-08-999-774A-2	Sequence 2, Appli
19	112.5	2.4	483	4	US-09-252-991A-31250	Sequence 31250, A
20	111.5	2.4	499	4	US-09-328-352-5415	Sequence 5415, Ap
21	111.5	2.4	1388	4	US-09-252-991A-20237	Sequence 20237, A
22	110.5	2.4	469	2	US-08-989-478-10	Sequence 10, Appl
23	110.5	2.4	469	3	US-08-996-685-10	Sequence 10, Appl
24	110.5	2.4	593	2	US-08-989-478-2	Sequence 2, Appli
25	110.5	2.4	593	2	US-08-989-478-8	Sequence 8, Appli
26	110.5	2.4	593	3	US-08-996-685-2	Sequence 2, Appli
27	110.5	2.4	593	3	US-08-996-685-8	Sequence 8, Appli
28	110.5	2.4	593	3	US-08-880-179-3	Sequence 3, Appli
29	110.5	2.4	859	4	US-09-199-637A-281	Sequence 281, App
30	110.5	2.4	859	4	US-09-252-991A-21413	Sequence 21413, A
31	110.5	2.4	1045	4	US-09-252-991A-30208	Sequence 30208, A
32	110.5	2.4	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
33	109.5	2.4	586	4	US-09-504-358-20	Sequence 20, Appl
34	109.5	2.4	586	4	US-09-954-314-20	Sequence 20, Appl
35	109	2.4	745	4	US-09-107-532A-5687	Sequence 5687, Ap
36	108.5	2.3	830	4	US-09-562-737-34	Sequence 34, Appl
37	108	2.3	421	4	US-09-585-173B-52	Sequence 52, Appl
38	107.5	2.3	751	1	US-08-123-702-4	Sequence 4, Appli
39	107.5	2.3	751	2	US-08-104-165-2	Sequence 2, Appli
40	107.5	2.3	751	2	US-08-422-333-2	Sequence 2, Appli
41	107.5	2.3	751	2	US-08-422-333-21	Sequence 21, Appl
42	107.5	2.3	751	3	US-08-464-250-2	Sequence 2, Appli
43	107.5	2.3	751	4	US-08-464-250-2	Sequence 2, Appli
44	107.5	2.3	751	4	US-08-832-867-5	Sequence 5, Appli
45	107.5	2.3	751	4	US-09-548-372D-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-09-604-605-2
; Sequence 2, Application US/09604605
; Patent No. 6421613
; GENERAL INFORMATION:
; APPLICANT: Ramgopal Nadimpalli
; APPLICANT: Carl R. Simmons
; TITLE OF INVENTION: Maize Prolifera Gene and Uses Thereof
; FILE REFERENCE: 1138
; CURRENT APPLICATION NUMBER: US/09/604,605
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/143,222
; PRIOR FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Zea mays
US-09-604-605-2

Query Match 16.7%; Score 774; DB 4; Length 720;
Best Local Similarity 32.4%; Pred. No. 4.7e-66;
Matches 200; Conservative 103; Mismatches 241; Indels 74; Gaps 13;

```
Qy        189 KNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFL---PEAPAELL 245
              |:||      ||   :   : |:      :: :  :||   : |   ||    |    :
Db         14 KDFLSNFTGPHGEPKYLNILQDVANRKIRAVQIELDDLFHYKDVDEEFLQRVTENTRRYI 73

Qy        246 QIFDEAALEVV-------------LAMYPKYD-----------------RITNHIHVR 273
               || ||  |::             : |  : |                 |     |
Db         74 GIFAEAMDEIMPEPTEAYTVDEDQDILMTQRVDEGADGGADGTDPLQKMPPEIKRFFEVY 133

Qy        274 ISHLPLVEEL--RSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNF-------- 323
              |     |  | | ::  :: ||:: ||:|| |: | | : :  | | :| |
Db        134 IKAFSKVTPLTIRQVKASNIGQLVKISGIVTRCSDVKPLMQVAVYTCEECGFEIYQEVTA 193

Qy        324 -VLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRS 382
               |  |  :  :|      |   :: |   : :  : :  :| :::||    |    :||:
Db        194 RVFMPLIECPSQR-----CKLNKAKGNLILQLRASKFLKFQEVKLQELAEHVPKGHIPRA 248

Qy        383 KDAILLADLVDSCNAGDEIELTGIY-HNNYDGSLNTANGFPVFATVILANHVAKKDNKVA 441
                |  :|      || :|::||:    | |      |  |  | : |  |     |
Db        249 LTVHLRGELTRKVAPGDVVEMSGIFLPMPYYGFRAMRAGL-VADTYLEAMSVTHFKKKYE 307

Qy        442 VGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHK 501
              :|  :: : |  |::|  |  |:  |:|| |:|||:|:   | | |   :      |
Db        308 EYDLKGDEQEQIDRLAEDGDIYSKLARSLAPEIFGHEDVKKALLLLLVGAPHRKLADGMK 367

Qy        502 VRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLE 561
```

```
Db        368 IRGDLHICMMGDPGVAKSQLLKHIINVAPRGVYTTGRGSSGVGLTAAVQKDPVTNEFVLE 427

Qy        562 AGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANP 621

Db        428 GGALVLADMGICAIDEFDKMEESDRTAIHEVMEQQTVSIAKAGITTSLNARTAILAAANP 487

Qy        622 IGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE 681

Db        488 AWGRYDMRRTPAENINLPPALLSRFDLLWLILDRADMETDLEMAR-----HVVHVHQNLE 542

Qy        682 EEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK 741

Db        543 ----------SPAL----GFTPLEPSVLRAYISAAR-RVIPSVPRELEEYIATAYSSIRQ 587

Qy        742 ESMATGSIPITVRHIESM 759

Db        588 EE-AKSNAPTSYTTIRTL 604


RESULT 2
US-09-198-452A-1072
; Sequence 1072, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1072
;   LENGTH: 918
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: 1...918
;   OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-1072

  Query Match             3.0%;  Score 140;  DB 4;  Length 918;
  Best Local Similarity   23.5%;  Pred. No. 0.0005;
  Matches   91;  Conservative   51;  Mismatches  141;  Indels  104;  Gaps   17;

Qy        494 KNPGGKHKVRGDI--NVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQR 551

Db        455 KNPNKFTSLGGRIPKGVLLIGPPGTGKTLIAKAVSGEADRPFF-------SIAGSDFV-- 505

Qy        552 HPVSREWTLEAGALVLAD-------RGVCL--IDEFDKM------------NDQDRTSIH 590

Db        506 -----EMFVGVGASRIRDMFEQAKRNAPCIIFIDEIDAVGRHRGAGIGGGHDEREQTLNQ 560

Qy        591 EAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILC 650

Db        561 LLVEMDGFGTNEGVILMAATNRPDVLDKALLRPGRFDRRVVMN-----LPDIKGRFEILM 615

Qy        651 V--VRDTVDPVQDEM-LARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQE 707

Db        616 VHAKRIKLDPTVDLMAVARSTPGA------SGADLENLLNEAALLAARKDRTAVTAVDVA 669

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGPRA 767

Db        670 EARDKVLYGKER-----------RSLEMDAEERK----------TTAYHES--------G 700

Qy        768 HPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNEL-------- 819

Db        701 HAVVGLC----VQHGDPVD-----KVTIIPRGLSLGAT-----HFLPEKNKLSYWKKELY 746

Qy        820 --LLFILKQLVAEQVTYQRNRFGAQQD 844

Db        747 DQLAVLMGGRAAEEIFLGDISSGAQQD 773


RESULT 3
US-09-328-352-6100
; Sequence 6100, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6100
;   LENGTH: 1515
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-6100

  Query Match             2.8%;  Score 128.5;  DB 4;  Length 1515;
  Best Local Similarity   19.4%;  Pred. No. 0.016;
  Matches  182;  Conservative  124;  Mismatches  328;  Indels  303;  Gaps   45;

Qy         22 SSRRTDALTSSPGRDLP--PFEDESEGLLGTEGPLEEEEDG--EELIGDGMERDYRAIPE 77

Db        281 TANRNWALARTPKFENPLLPGLTELNPIVNRTGSDSSSLDNMLEILVGGGMDL-FRALRM 339

Qy         78 L--DAYE-AEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCM---TAMR 131

Db        340 LVPPAWQNVETLDADLRAFYEFNSKHMEAWDGPAGLVIQD----GRHAI--CMLDRNGLR 393

Qy        132 RTRSALP----ASAASGAGTEDGEEDEQMIE----------------SIENLEDLKGH- 169

Db        394 PARWVITKNDYITLASEIGVW-GYEPEDVVSKGRVGPGQILVVDTLTGKVLDTKDVSNHL 452

Qy        170 ----SVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 225

Db        453 KNMRPYREWLRDHAIRLNANPELEEQL----------VDKGLTGDALKAAQKMFMVTFEE 502

Qy        226 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHL--PLVEEL 283

Db        503 ---RDQLLRPIAESGQEAVGSMGDDTPMAV---LSRQVRHVTDYFRQQFAQVTNPPIDPL 556

Qy        284 R-----SLRQL------------HLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVL 325

Db        557 RESIVMSLETCLGREQNVFEQGPEHADRIIISSPVLSN---------------------- 594

Qy        326 GPFCQSQNQEVKPGSCPECQSAG----PFEVNMEETIYQNYQRIRIQESPGKVAARRLPR 381

Db        595 -----SKMQQIR-----SIERAGYEAVDIDLNYAETEGLQAAITRICEE----SAQAVRD 640

Qy        382 SKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVA 441

Db        641 GKTLIVLTD----------------KNIRQGYL-PANA--ALATGAVHHHLIK------ 674

Qy        442 VGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHK 501

Db        675 TGLRTDANIMVETGFARDPHQFAVLLGFGATAIY--------PYLA------------YD 714

Qy        502 VRGDINVL--LCGDPGTAKSQFLKYIEK----------VSSRAIFTTG------------ 537

Db        715 VINDLIAKGELLGDPIHAQANFRKGIEKGLLKVLSKMGISTVASYRGGQLFEAVGLSSEV 774

Qy        538 ------------QGASAV---------AVTAYVQRHPVSREWTLE--------------- 561

Db        775 VDQCFLGVPSRIQGATFVDLENDQKKLAATAWSNRKPIDQGGLLKFVFGKEYHAFNPDVI 834

Qy        562 --------AG--------ALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSI--SISKA 603

Db        835 NSLHKAVRSGKYEDFKEYAELVNNRPIATIRDLFKLKTTNPIPVEQVESVEAILPRFDSA 894
```

```
Qy        604 GI---VTSLQARCTVIAAANPIGGR-------YDPSLTFS-ENVDLTEPIISRFDILCVV 652
              |:     | :|   :  | | ||||        ||:   :  |  : :    || :
Db        895 GMSLGALSPEAHEAIAIAMNTIGGRSNSGEGGEDPARYGTIRNSKIKQIASGRFGVTPAY 954

Qy        653 RDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAE------------PAMPNTY 699
              : : :|        |    :     :   |  ||  |              |   : |
Db        955 LTSAEVLQ------IKVAQGAKPGEGGQLPGGKVNGLIARLRYSVPGVTLISPPPHHDIY 1008

Qy        700 GVEPLPQEVLKKYIIYAKERVHPKLNQMDQDK------------VAKMYSDLRKESMAT 746
               :| | |      :|:  ::|:||   |  |             ||| |:|    |
Db       1009 SIEDLSQ------LIFDLKQVNPK--AMVSVKLVSEPGVGTIAAGVAKAYADFITISGYD 1060

Qy        747 GSIPITVRHIESMSHGGGP------RAHPSAGLCDRR 777
              |       : |: | | |       || :  : | |
Db       1061 GG--TAASPLSSIHHAGSPWELGLSEAHQALRVNDLR 1095


RESULT 4
US-09-134-001C-3979
; Sequence 3979, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3979
;   LENGTH: 871
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3979

  Query Match             2.7%;  Score 126.5;  DB 4;  Length 871;
  Best Local Similarity   18.4%;  Pred. No. 0.0094;
  Matches  164;  Conservative  150;  Mismatches  285;  Indels  293;  Gaps   45;

Qy         23 SRRTDALTSSPGRDLPPFEDE---SEGLLGTEGPLEEEEDGEELIGDGMERDYRAI---- 75
              | :|: |     : :  :|||    | :|     :: :|  :: :|: :|     |
Db         88 SAKTNELLDKAEKYMKSYEDEFISMEHIL--RAAIDTDETTQKWVGNKVEVIKEIITKVR 145

Qy         76 ---------PELDAYEA-------------EG-----LALDDEDVEELTASRREAADGPC 108
                      ||:: |||             :|     :  |:|   :     |:  : |
Db        146 GGNHVTSQNPEVN-YEALEKYGRDLVEEVRQGKMDPVIGRDEEIRNTIRILSRKTKNNP- 203

Qy        109 GTVTGSWPGLGACAVGSCMTAMRRTRSALPAS-----------AASGAGTE-DGEEDEQM 156
                |    ||:|  |:   : | |  |  :| |           :|  || :  || :|::
Db        204 --VLIGEPGVGKTAIVEGL-AQRIVRKDVPESLLDKTIFELDLSALVAGAKYRGEFEERL 260

Qy        157 IESIENLEDLKGHSV-----------------REWVSMAGPRL---EIH-------HRF 188
                 :: ::: :|  :                  :  :|  | |   |:|       : :
Db        261 KAVLKEVKESEGRIILFIDEIHMLVGAGKTDGAMDAGNMLKPMLARGELHCIGATTLNEY 320

Qy        189 KNFLRTHVDSHGHNVF-KERISDMCKENRESLV-----------VNYED---LAARE-- 230
              : ::    ||     | |  :|:   ||  |::           |  :|   :|| |
Db        321 REYIEK--DSALERRFQKVGVSEPDVENTISILRGLKERYEVYHGVRIQDRALVAAAELS 378

Qy        231 --HVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVE------- 281
                ::   |||:   :|:   |:|   :   |      : : :: |:  | : |
Db        379 DRYITDRFLPDKAIDLV---DQACATIRTEMGSNPTEL-DQVNRRVMQLEIEESALKNES 434

Qy        282 ------ELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQE 335
                   |  |::   |:  : | :           | |:    |         : | :
Db        435 DNASKHRLEELQEELSNEKEKQSSL----------KSRVEQEKEK-------IAKVQEKR 477

Qy        336 VKPGSCPECQSAGPFEVNMEETIYQNYQRI-----------RIQESPGKVAARRLPRSK 383
               :  |  :       | |:|:     |  |             |:  |: : |
```

```
Db        478 AELDSSRQALEDAQTEGNLEKAAELQYGTIPQLEKELQEFEEAFQDETGEDSER------ 531

Qy        384 DAILLADLVDSCNAGDEI-ELTGIYHNNYDGSLNTANGFPVFATV------------ILA 430
                 :: ::|      || : : |||              ||   |            ||
Db        532 ---MIREVVSDEEIGDIVSQWTGI--------------PVSKLVETEREKLLSLSDILH 573

Qy        431 NHVAKKDNKVAVGELTDEDVKMITSLSKD--QQIGEKIFASIAPSIYGHEDIKRGPALAL 488
                |  :|  |   :|  : |    :  ||  : ||  :|  : |:  |  :: :  | :|
Db        574 KRVVGQDKAV---DLVSDAVVRARAGIKDPNRPIGSFLF--LGPTGVGKTELAKSLASSL 628

Qy        489 FGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQG----ASAVA 544
              |  |      || :| |::                :|:|| :   :     |
Db        629 FDSE------KHMIRIDMS--------------EYMEKHAVSRLIGAPPGYVGHDEGGQ 667

Qy        545 VTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAG 604
              :|  |:|:| |                  | |:|| :| :       : : :::  :: ||
Db        668 LTEAVRRNPYS---------------VILLDEVEKAHSDVFNVLLQILDEGRLTDSKG- 710

Qy        605 IVTSLQARCTVIAAANPIGGRYDPSLTFSENV-------DLTE------------PIIS 644
                 |:  : |:|   : ||     |    |||       | ||              |::
Db        711 --RSVDFKNTIIIMTSNIG-----SQVLLENVKDAGEISDDTEKAVMDSLHAYFKPEILN 763

Qy        645 RFDILCVVR----DTVDPVQDEMLARF---VVGSHVRHHPSNKEEEGLANGSAAEPAMPN 697
              | | : : :    | :  : |::| :    ::  |:    : : :: |      | |
Db        764 RMDDIVLFKPLSVDDMSMIVDKILTQLNMRLLDQHISIEVTEEAKKWL-----GEEAYEP 818

Qy        698 TYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSI 749
               :|  |     ||:::           :  :  :|:|   : |||:  |:|
Db        819 QFGARP-----LKRFV----------QRQIETPIARM---MIKESLPEGTI 851


RESULT 5
US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
;  GENERAL INFORMATION:
;    APPLICANT:  Leitner, Ernst
;    APPLICANT:  Schneider, Elisabeth
;    APPLICANT:  Schoergendorfer, Kurt
;    APPLICANT:  Weber, Gerhard
;    TITLE OF INVENTION:  Cyclosporin Synthetase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5827706artis Corporation
;      STREET:  59 Route 10
;      CITY:  East Hanover
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07936
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,119A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kassenoff, Melvyn
;      REGISTRATION NUMBER:  26,389
;      REFERENCE/DOCKET NUMBER:  100-8029/CONT/CONT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 503 8474
;      TELEFAX:  201 503 8807
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15281 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Tolypocladium niveum
;      STRAIN:  ATCC 34921
US-08-471-119A-2

  Query Match             2.7%;  Score 125;  DB 2;  Length 15281;
  Best Local Similarity   18.3%;  Pred. No. 2.2;
  Matches  185;  Conservative  124;  Mismatches  331;  Indels  372;  Gaps   47;

Qy          7 QRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTE---GPLEEEEDGEEL 63
              |:  | ||  |:            :   : ||    ::  |:||    | |:: |   ||
Db       9156 QKYVGLDPAPSAA--------IFVNEAVKSLPSLAGKARVLVGTALDIGSLDKNEIQPEL 9207

Qy         64 I---------------------------------GD----GMERDY---RAIPEL-DAYE 82
              :                                 ||     : ||:   ||:  | |
Db       9208 VVINSVAQYFPTSEYLIKVVKAVVEVPSVKRVFFGDIRSQALNRDFLAARAVRALGDNAS 9267

Qy         83 AEGLALDDEDVEELTASRREAADGPCGTVT--GSWPGLGACAVGSCMTAMRRTRSALPAS 140
              | :    | : ||  |  |     |   |:     | :    |      |: |
Db       9268 KEQIR---EKIAELEESEEELLVDPAFFVSLRSQLPNIKHVEV--LPKLMKATNELSSYR 9322

Qy        141 AASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHG 200
              |:       ||:: :|: |:            ||               :|  |  || |
Db       9323 YAAVLHISHNEEEQLLIQDID---------PTAWV-------------DFAATQKDSQG 9359

Qy        201 HNVFKERISDMCKENRESLV-----VNYEDLAAREHVL---------------------- 233
                    : :: :: |: ::     : |       |::
Db       9360 -------LRNLLQQGRDDVMIAVGNIPYSKTIVERHIMNSLDQDHVNSLDGTSWISDARS 9412

Qy        234 -----AYFLPEAPAELLQIFDEAALEVVL----------AMYPKYDRITNHIHVRIS--- 275
                  |  :||| | |:  |    | |              |:   : |:    :   |
Db       9413 AAAICTSF--DAPA-LTQLAKEEGFRVELSWARQRSQNGALDAVFHRLATDANCERSRVL 9469

Qy        276 -HLPLVEELRSLRQLHLNQLIRTSG----------------VVTSCTGVLPQLSMVKY 316
               | |   : | ||  |    | |                       :: |   ||||:
Db       9470 VHFPTDHQGRQLRTLTNRPLQRAQSRRIESQVFEALQTALPAYMIPSRIIVLPQM----- 9524

Qy        317 NCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAA 376
                                          |   | :   |  :| ::     |:|
Db       9525 -----------------------------PTNANGKVDRKQLARRAQVVAKRKAVSA 9552

Qy        377 RRLPRSKDAILL----ADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANH 432
              |  ||:   |:|    ||::     | |:   ||  | :|       |  : || : |
Db       9553 RVAPRNDTEIVLCEEYADIL-----GTEV---GITDNFFD-----MGGHSLMATKLAARL 9599

Qy        433 VAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
               : | :| | |: |: |  :  |:   : |      || |:|       ||       |
Db       9600 SRRLDTRVTVKEVFDKPV--LADLAASIEQGSTPHLPIASSVYS------GPV------E 9645

Qy        493 PKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRH 552
                 |:       |:                 |   :: | :      |  ||: |  |||
Db       9646 QSYAQGRLWFLDQFNL-----------NATWYHMSLAMRLLGPLNMDALDVALRALEQRH 9694

Qy        553 PVSREWTLEA----GALVLADRG---VCLIDEFDKMNDQDRTSIHEAMEQQSISI-SKAG 604
                 | | ||    |  |: : |   : ::|  || |:::  :: |  : :  :: |: |
Db       9695 ETLRT-TFEAQKDIGVQVVHEAGMKRLKVLDLSDK-NEKEHMAVLENEQMRPFTLASEPG 9752

Qy        605 IVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEML 664
                 |               |  |           || |:|
Db       9753 WKGHL-------------ARLGP----------TEYILSL------------------- 9769

Qy        665 ARFVVGSHVRHHPSN--------KEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYA 716
                    | ||  :        ::| |    :|     | :  |:||| :  :  :  :
Db       9770 --------VMHHMFSDGWSVDILRQELGQFYSAALRGRDPLSQ-VKPLPIQ-YRDFAAWQ 9819

Qy        717 KERV----HPKL-----NQMDQDKVAKMYSDLRKESMATGS---IPITVRHIESMSHGGG 764
              ||       | :      ||:      :: :|  :    :|    ||:|:           |
Db       9820 KEAAQVAEHERQLAYWENQLADSTPGELLTDFPRPQFLSGKAGVIPVTIE---------G 9870

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFIL 824
              |                                      : : | | ::    |  :|
Db       9871 P---------------------------------------VYEKLLKFSKERQVTLFSVL 9891

Qy        825 KQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRM 876
                    | | : |: |   ||:  ||  |   : :: |    | :  | :::  |:
Db       9892 --LTAFRATHFR-LTGAEDATIGTP---IANRNRPELEHIIGFFVNTQCMRL 9937


RESULT 6
US-09-252-991A-24044
; Sequence 24044, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 24044
;   LENGTH: 516
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24044

  Query Match             2.6%;  Score 121.5;  DB 4;  Length 516;
  Best Local Similarity   23.0%;  Pred. No. 0.011;
  Matches   66;  Conservative   43;  Mismatches  109;  Indels   69;  Gaps   11;

Qy        504 GDINVLLCGDPGTAKS-------QFLKYIEKVSSRAIFTTGQGASAVAVTAYVQR----- 551
                |  |:|| | ||| |:          |  :::  :  :      || | :  : ||
Db        229 GAHNLLLSGPPGTGKTLLASRLPGLLPALDEDEALEVAAIHSVASHVPLRHWPQRPFRQP 288

Qy        552 -HPVSREWTL------EAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAG 604
                | |    :      : | : || :||  :|| :   :    : | :|   | |::|
Db        289 HHSASAPALVGGGSRPQPGEITLAHQGVLFLDELPEFERKVLEVLREPLESGEIVIARAN 348

Qy        605 IVTSLQARCTVIAAANPIGGRY--DPS---LTFSENVD-----LTEPIISRFDI-LCVVR 653
                    ||  ::|| ||    |  |||       | |      |: |:: | |: : |:|
Db        349 GRVRFPARFQLVAAMNPCPCGYLGDPSGRCRCTPEQVQRYRGKLSGPLLDRIDLHVSVLR 408

Qy        654 DT-----------------VDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAM 695
              ::                    |   :  |||                :| ||      ||
Db        409 ESTSLQPGHGETATAEVSERVGAARQRQLAR---------------QGCANAHLDLQAM 452

Qy        696 PNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQD-KVAKMYSDLRK 741
                    :     |  ::::  | ||:   |  : :  |||:  :|| :
Db        453 HRNCAL----AEADRRWLEAAGERLELSLRALHRILKVARTLADLER 495


RESULT 7
US-09-107-532A-4537
; Sequence 4537, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
```

```
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 4537:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 875 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
;            ORGANISM: Enterococcus faecium
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...875
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4537:
US-09-107-532A-4537

  Query Match             2.6%;  Score 121;  DB 4;  Length 875;
  Best Local Similarity   18.0%;  Pred. No. 0.032;
  Matches  149;  Conservative  135;  Mismatches  313;  Indels  232;  Gaps   39;

Qy         54 LEEEEDGEELIGDGMERDYRA--------IPELDAYEAEGLALDDEDVEE-LTASRREAA 104
              :||   || :     |  |:|        : :: : : : :   ||:: : :    |:
Db        148 IEEMRGGERVTSQNQEEQYKALEKYGVDLVQQVKSGKMDPIIGRDEEIRDVIRILSRKTK 207

Qy        105 DGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLE 164
              : |   |    ||:|  |:   : | |  |  :|                      |||:
Db        208 NNP---VLIGEPGVGKTAIVEGL-AQRIVRKDVP----------------------ENLK 241

Qy        165 DLKGHSVREWVSMAGPRL--EIHHRFKNFLRTHVDSHGH-NVFKERISDMCKENRESLVV 221
              |    |:     :|| :   |   | |  |:    | |   :| : | ::    :
Db        242 DKTIFSLDMGALIAGAKFRGEFEERLKAVLKEVKKSDGRILLFIDEIHNIVGAGKT---- 297

Qy        222 NYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHL---- 277
              |      ::|   |      |:          |  | :|      :  |   :
Db        298 --EGSMDAGNLLKPMLARGELHLIG-------ATTLDEYRQYMEKDKALERRFQKVLVKE 348

Qy        278 PLVEE----LRSLR---QLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 330
              | ||:    || |:   ::|    |  : :| : |     ||          ::   |
Db        349 PTVEDTISILRGLKERFEIHHGVNIHDNALVAAAT-----LS--------DRYITDRFLP 395

Qy        331 SQNQEVKPGSCPECQSAGPFEVNMEET----IYQNYQRIRIQESPGK-----VAARRLPR 381
              :  ::       |  :    |:|   |    : :   :: |:|:  |       : :||
Db        396 DKAIDL----IDEASATIRVEMNSMPTELDQVTRRLMQLEIEEAALKKESDDASKKRLKN 451

Qy        382 SKDAILLADLVDSCNA---------------GDEIELTGIYH------NNYD---GSLN 416
              ::   ||:| :  ||                  |:    |      ||||    ::
Db        452 LQEE--LAELREEANAMKMQWETEKEEVNSVSAKRAEIDKAKHELEDAENNYDLERAAVL 509

Qy        417 TANGFPVFATVILANHVAKKDNKV--------------AVGELTDEDVKMITSLSKDQQI 462
                  |    :       ||:::              || ||   |  :    ::: |
Db        510 RHGTIPQLEKELKELEAKAKDSEIKMVQESVTENEIAQVVGRLTGIPVTKLVEGEREKLI 569

Qy        463 GEKIFASIAPSIYGHEDIKRGPALALFGG-----EPKNPGGKHKVRGDINVLLCGDPGTA 517
              |:  ::   : | ::     : |:        :|  | |         : |  |  |
Db        570 --KLNETLHKRVIGQDEAVDAVSDAVIRSRAGLQDPNRPLG--------SFLFLGPTGVG 619

Qy        518 KSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSR-----------EWTLEAGALV 566
              |::    : |  :  :| :      : :: |:::| |||           |   | | |

Db        620 KTE----LAKALAENLFDSEDHMVRIDMSEYMEKHAVSRLVGAPPGYVGYE---EGGQLT 672

Qy        567 LADR----GVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPI 622
              | |     : |:|| :| :       : : ::   :: ||  :|     : ||:   : |
Db        673 EAVRRNPYTIVLLDEIEKAHPDVFNILLQVLDDGRLTDSKGRVV---DFKNTVLIMTSNI 729

Qy        623 GGRYDPSLTFSENVDLTEP-----------------IISRFDILCVVR----DTVDPVQ 660
              | :    |      | | |                    ::| |   :      | |  :
Db        730 GSQL---LLEGVTADGTIPEAVAEQVNTLLRGNFKPEFLNRIDDTILFTPLSLDNVKGIV 786

Qy        661 DEMLARFVVGSHVRHHP-----SNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIY 715
              |:|:|:  :   : |       |:: :  :|  :| |||    ||  || : : |
Db        787 DKMVAQ--LAQRLEHQEILLTISDEAKTWIAE-NAYEPA----YGARPLKRFITK----- 834

Qy        716 AKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIP---ITVRHIESMSH 761
                                  :: : | || :|   :|   :|:  ::   |
Db        835 -----------------EVETPLAKEIVAGHVMPKSKVTITLLDGQLH 865


RESULT 8
US-09-134-001C-4008
; Sequence 4008, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4008
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4008

  Query Match             2.6%;  Score 121;  DB 4;  Length 1162;
  Best Local Similarity   20.0%;  Pred. No. 0.054;
  Matches  174;  Conservative  107;  Mismatches  313;  Indels  274;  Gaps   45;

Qy        183 EIHHRFKNFLRT---HVDSHGHNVFKE-------RISDMCKENRESLVVNY-----EDLA 227
              |:|      |||    | :   | ||:|       |:|   |:  | :: :       ||
Db         59 ELH----GSLRTEVLHFERLSHRVFQEVGGLTEQRLS---KDALEMMIFHIVQQHESDLK 111

Qy        228 AREHVLAYF-LPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSL 286
                    |: | |  || :| |             ||                      ::
Db        112 LYGSQAQYYGLSEKLAEQIQDF-----------KKY--------------------NV 138

Qy        287 RQLHLNQLIRTSGVVTSCTGVLPQLSMV-KYNCNKCN--FV-----LGPFCQ--SQNQEV 336
                  ||||||    : |     |  :|:: |    ::  |  |:       |  | :  ||:| :
Db        139 TPEHLNQLIENHSIQTRTKHKLEDISLIYKQLESRMNGEFITTEDSLQQFIEILSQSQWI 198

Qy        337 KPGS--CPECQSAGPFEVNMEETIYQNYQRIRI-QESPGKVAARRLPRSKDAIL--LADL 391
              |          :    |  : | : |: ::: :   :  | |      | |    :|  | |:
Db        199 KKAEVFIDGFHNFSTLEYRIIEALVQHAKQVTVLLTTDGSHHPFSLFRKPSEVLSHLEDI 258

Qy        392 VDSCNAGDEIELTGIYHN-----NYDGSLNTANGFPV--FATVILANHV--------AKK 436
              :  |    | |   | |      | |   |  |||     |       |||          ::
Db        259 ANRLN----INLNKTYFNTFYRYNNDDLKNLENGFDALQFTPKHHQNHVKIFESSSMREE 314

Qy        437 DNKVA---VGELTDEDVKM--ITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGG 491
               |:||   : :: : | |    |  | :|:     :| || ||     ||        |
Db        315 INEVARRILKDVREADYKFRDIAILYRDESYA-YLFESILPSY----DIP-------FNI 362

Qy        492 EPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSR-------AIFTTGQGASAVA 544
              : |     | :   :  ||                  | : |          :| |    |
Db        363 DTKKSMTHHPIMEMLRSLL--------------EVIRSNWHINAMLRLFKTNVLTSQFK 407
```

```
Qy        545 VTAY---------VQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDR--TSIHEAM 593
              ::|         ::|    : |          |  :  ||:| :|  :    |  |:|:
Db        408 RSSYLIDLLENFVLERGIYGKRW---------LDEDIFSIDQFSRMGRKSHQLTEGHQAL 458

Qy        594 EQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVD---------------- 637
              :: | : |  |   |        | |       | : :| |:::
Db        459 YKEVIKLKKNVINKVLYFE----QAMNEAHTVKDYATSFYESLEYFELPSQLMTQRDELE 514

Qy        638 ---LTEP----------IISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHH---PSNKE 681
                 |||           :|   | |  | |  :    : |  | :|         |   :
Db        515 LAGLTEKAEEIDQVWNGLIQILDDLVTVFDDQEMTLQQFLDVFDIGLEQLEFVMIPQTLD 574

Qy        682 EEGLANGSAA--------------EPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLN-- 725
              :  :     |            |  :  |||:   ::  | :|:
Db        575 QVSIGTMDLAKVDNKKHIYMVGMNDGILPQTVSSSSLITDEEKKYV---EDNAHVELSPT 631

Qy        726 ----QMD------------QDKVAKMYS------DLRKESMATGSIPITVRHIESMSHGG 763
                  |||            |  |   ||      | :: |     |      :| ::
Db        632 SDILQMDEAFVCYIAMTRSQQSVTFSYSLMGNSGDEKEISPFLTQIKELFYDLEIINLQD 691

Qy        764 GPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMR----------KTFARYLSFR 813
              :| |   :      || |          |::|    :|          :     ||: |
Db        692 LHKAQPLLMM------QHSH---------QTKIQLFEYLRGWLDHEDIDYRWLDAYLAIR 736

Qy        814 RDN--NELLLFILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAF--- 868
              |:  |: | ::   |  :  | | |   :||         | |    ::      :
Db        737 DDDQLNQGLDYLTTSLTYDNETVQLNEILSQQ---------LYGKTINASVSRFEGYQQC 787

Qy        869 ----YDSELFRMNKFSHDLKRKMILQQF 892
                  | |   |:|:     : |  || |
Db        788 PFKHYASHGLRLNE-----RTKYELQNF 810


RESULT 9
US-08-755-587-189
; Sequence 189, Application US/08755587
; Patent No. 6045997
;  GENERAL INFORMATION:
;    APPLICANT:  Futreal, Phillip A
;    APPLICANT:  Wooster, Richard F
;    APPLICANT:  Ashworth, Alan
;    APPLICANT:  Stratton, Michael R
;    TITLE OF INVENTION:  Materials and methods relating to the
;    TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
;    TITLE OF INVENTION:  susceptibility gene and uses thereof.
;    NUMBER OF SEQUENCES:  222
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson
;      STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;      CITY:  Raleigh
;      STATE:  NC
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/755,587
;      FILING DATE:  25-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9523959.6
;      FILING DATE:  23-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9525555.0
;      FILING DATE:  14-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9617961.9
;      FILING DATE:  28-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kenneth D Sibley
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5405-135
;  INFORMATION FOR SEQ ID NO:  189:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1589 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-755-587-189

  Query Match              2.6%;  Score 118.5;  DB 3;  Length 1589;
  Best Local Similarity   18.1%;  Pred. No. 0.17;
  Matches  144;  Conservative  123;  Mismatches  329;  Indels  199;  Gaps   35;

Qy         21 RSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPE-LD 79
              ::::  |: |||   :|   |:     :  | ||:   :   :|  |   :     ||
Db        664 KNTKHEDSYTSSQRNNL---ENSDGSMSSTSGPVYIHKGDSDLPADQGSK----CPESCT 716

Qy         80 AYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPA 139
               |  |      |:: :||    :| :                        :|| :    :  ||:
Db        717 QYAREENTQIKENISDLTCLIMKAEE------------------TCMKS--SDKKQLPS 755

Qy        140 SAASGAGTEDGEEDEQMIESIE-NLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDS 198
                      :: || |:    : :   | : |          : | :|  : |     ||
Db        756 -----------DKMEQNIKEFNISFQTASGKNTRVSKESLNKSVNIFNRETDELTVISDS 804

Qy        199 HG----HNVFKERISDMCKENRESLVVNYED--------LAAREHVLAYFLPEAPAELLQ 246
                    | : |:::   | :   |:   :||         | |::|    | :        |
Db        805 LNSKILHGINKDKMHTSCHKKAISIKKVFEDHPIVTVSQLPAQQHP-EYEIESTKEPTLL 863

Qy        247 IFDEAALEVVLAMYPKYDRITNHI----HVRIS------HLPLVEELRSLR------QLH 290
              |  |: : |  |     |:: |      :|| :          || :  : |      ::
Db        864 SFHTASGKKVKIMQESLDKVKNLFDETQYVRKTASFSQGSKPLKDSKKELTLAYEKIEVT 923

Qy        291 LNQLIRTSGVVTSCTGVLPQLSMVKY----NCNKCNFVLGPFCQS--QNQEVKPGSCPEC 344
               ::       |:  | :|||  :   |    |    |       ::   | |  |  |  |
Db        924 ASKCEEMQNFVSKETEMLPQQNYHMYRQTENLKTSNGTSSKVQENIENNVEKNPRICCIC 983

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
              ||: |     |::      |    |    |    ||     | :  |      : |::
Db        984 QSSYPV---TEDSALAYY-----TEDSRKTCVRESSLSKGRKWLREQGDKLGTRNTIKIE 1035

Qy        405 GIYH-NNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIG 463
                |  :: |: :  :   :  | |  |||::                  :::|  |  :
Db       1036 CKEHTEDFAGNASYEHSLVIIRTEIDTNHVSE---------------NQVSTLLSDPNV- 1079

Qy        464 EKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGG---KHKVRGDINVLLCGDPGTAKS- 519
                 : :: | : | |            :  |  |   |:|:  |:   :    |   |
Db       1080 --CHSYLSQSSFCHCD-----------DMHNDSGYFLKNKIDSDVPPDMKNAEGNTISP 1125

Qy        520 ------------QFLKY-IEKVSSRAIFTTGQGASAVAV-------------TAYVQRH 552
                         | : | ::|: :    |:      ||:                ::  |
Db       1126 RVSATKERNLHPQTINYCVQKLETN---TSPHANKDVAIDPSLLDSRNCKVGSLVFITAH 1182

Qy        553 PVSREWTLEAGALVLADRGVCLIDEFDKMN-DQDRTSIHEAMEQQSISI-SKAG------ 604
              | | |    :: |    ::::  :   |  :|| |: ::     | | :|
Db       1183 SQETERTKE----IVTDNCYKIVEQNRQSKPDTCQTSCHKVLDDSKDFICSSSGDVCINS 1238

Qy        605 -----------IVTSLQARCTVIAAANPIGG--RYDPSLTFSENVDLTEPIISR-FDILC 650
                        |:   |:   :  || |  |   :| | :  |      |  || : |
Db       1239 RKDSFCPHNEQILQHNQSMSGLKKAATPPVGLETWDTSKSIREPPQAAHP--SRTYGIFS 1296

Qy        651 VVRDTVDPVQD----------------EMLARFVVGSHVRHHPSNKEEEGLANGSAAEP 693
                    | |                : |:  |:  : : | | | |  : : :
Db       1297 TASGKAIQVSDASLEKARQVFSEMDGDAKQLSSMVLEGNEKPHHSVKRENSVVHSTQGVL 1356

Qy        694 AMPNTYGVEPLPQEV 708
              ::|     :|||  |
Db       1357 SLP-----KPLPGNV 1366


RESULT 10
US-09-315-793-12
; Sequence 12, Application US/09315793
; Patent No. 6221597
; GENERAL INFORMATION:
```

```
;  APPLICANT: Roberts, Christopher J.
;  TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
;  TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
;  TITLE OF INVENTION: DRUGS
;  FILE REFERENCE: 9301-048
;  CURRENT APPLICATION NUMBER: US/09/315,793
;  CURRENT FILING DATE: 1999-05-21
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 1698
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-315-793-12

  Query Match             2.5%;  Score 117.5;  DB 3;  Length 1698;
  Best Local Similarity   19.0%;  Pred. No. 0.23;
  Matches  117;  Conservative   80;  Mismatches  196;  Indels  223;  Gaps   27;

Qy        220 VVNYEDLAAREHVLAYF------LPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVR 273
              :: |  :: | |::  :      ||:    ||         :: :: |  |  :|
Db        125 LMTYASMSVRSHLIELYDNYILLLPQTTLRLL------IRPLISSLLPGIDDESN----- 173

Qy        274 ISHLPLVEELRSLRQLHLNQ---------LIRTS------GVVTSCTGVLPQLSMVKYNC 318
                 |||  :|    | :|:          |: |:      | :|  |   | |: | :
Db        174 -DFLPLTLKLIETLQENLDDDSLFWQTLFLVMTANKGRRLGGLTWLTRKFPSLNAVPHLV 232

Qy        319 NKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKV---A 375
              ||                |  |                 ||   :  | : :|   ||   |
Db        233 NKI---------KMEAEENPSE-------------TETNDSHLDRKKRKEEAFKVLLPA 269

Qy        376 ARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVI-LANHVA 434
              |: |   :  :|:  ||                    | |   |   :  :|: |
Db        270 AKDLVTPEPGLLIRCLV-------------------GCLEDENDILIKRSVLDLLLQRL 309

Qy        435 KKDNKVAVGELTDEDVKMI------TSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALAL 488
              : |: |    :| || |::      |:|||| :  :|:  :                ||
Db        310 RLDSPVLNVLITSEDKKLLIMSCCRTTLSKDMSLNRRIWNWLL-----------GPTA-- 356

Qy        489 FGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSS----RAIFTTGQGASAV- 543
               ||   | ||                        ::|   | |      :: |  | ||:
Db        357 -GGMLNNNGGNS----------,---------MEYTTSVKSANEESNVYFTKYGLSALL 394

Qy        544 ------------AVTAYVQRHPVSREWTLEAGALVLADRGVCLI--DEFDKMNDQDRTSI 589
                         :||:     |   |  | |:||: :  : |:   |  | |:|   :
Db        395 EGLSDLLSEEESVLTAFRISMAVMDRW--EIGSLVIPELFIPLLYSSEKFKQNEQIMKTA 452

Qy        590 HEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDIL 649
                  :  :|                | |:    |   :|: : : ::: |:|
Db        453 RTFFDNTETNI--------------------IWGKLFQELEDIKNLKILDFVLTNFNI- 490

Qy        650 CVVRDTVDPVQDEMLARFVVGSHVRH----------HPSNKEEEGLANGSAAEPAMPNTY 699
                       || :        |||           ||| ::               | |
Db        491 ---------GNDEEII-------VRHLPLILLTLLALPSNDKD-----------FDNIY 522

Qy        700 GVEPLPQEVLKKYIIYAKERV-----HPKLNQMDQ-------DKVAKMYSD-------LR 740
               ::     :  | : |  ||      | ||   |:        |:   |::       |
Db        523 KLQKF--SLYNKLLNYIPERALLPLSHSKLKHDDEVSCEELLAKIRGFYTNVSNPSSILE 580

Qy        741 KESMATGSIPITVRHI 756
              ||::|    | |   :
Db        581 KENIAERLPPFTTEDL 596


RESULT 11
US-09-001-826-5
; Sequence 5, Application US/09001826A
; Patent No. 5965727
; GENERAL INFORMATION:
;  APPLICANT: SONG, HEE-SOOK
;  APPLICANT: BROTHERTON, JEFFREY E.
;  APPLICANT: WIDHOLM, JACK M.
;  TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
;  TITLE OF INVENTION: TRANSFORMATION
;  FILE REFERENCE: UI001.C1
;  CURRENT APPLICATION NUMBER: US/09/001,826A
;  CURRENT FILING DATE: 1997-12-31
;  EARLIER APPLICATION NUMBER: 08/937,739;  60/025,140
;  EARLIER FILING DATE: 1997-07-25; 1996-07-26
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text
; SEQ ID NO 5
;   LENGTH: 616
;   TYPE: PRT
;   ORGANISM: Nicotiana tabacum
US-09-001-826-5

  Query Match             2.5%;  Score 117;  DB 2;  Length 616;
  Best Local Similarity   21.0%;  Pred. No. 0.043;
  Matches  137;  Conservative   84;  Mismatches  231;  Indels  200;  Gaps   37;

Qy        148 EDGEEDEQMIESIENLEDLKGHSV-REWVSMAGPRLEI----------HHRFKNFLRTH 195
              :| |    : ||:|     :| || |  |  | | :||          ||  |   :|
Db         91 DDREAPSFLFESVE--PGFRGSSVGRYSVVGAQPSMEIVAKEHNVTILDHHTGKLTQKTV 148

Qy        196 VD--------SHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLP--EAPAELL 245
               |        | |   :| |: |   :      | |       :|    ||   || :
Db        149 QDPMTIPRSISEG---WKPRLIDELPDTFCGGWVGYFSYDTVRYVENRKLPFLRAPED-- 203

Qy        246 QIFDEAALEVVLAMYPK---YDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVT 302
               |    :: | :|     :| :   ||   |  :::  || : :|:  |   :|:
Db        204 ---DRNLADIQLGLYEDVIVFDHVEKKAHV--IHWVQLDQYSSLPEAYLDGKKRLEILVS 258

Qy        303 SCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSA-GP--FEVNMEETIY 359
              |:                          ::  : |||   |  | ||   : ||    |
Db        259 RVQGI----------------------ESPRLSPGSVDFCTHAFGPSLTKGNMTSEEY 294

Qy        360 QNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTAN 419
              :|     : ::   :||  : :    |:|:   :      |  |   :|      :|   |
Db        295 KN----AVLQAKEHIAAGDIFQ----IVLSQRFERRTFADPFE---VYR-----ALRIVN 338

Qy        420 GFPVFATVILAN---------HVAKKDNKVAV-----------GELTDEDVKMITSLSKD 459
               | : | | |            :  :  |  :           |:  |||| :   : ||
Db        339 PSP-YMTYIQARGCILVASSPEILTRVKKRRIVNRPLAGTSRRGKTPDEDVMLEMQMLKD 397

Qy        460 Q-QIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINV--LLCGD--- 513
              : |  | |                    |   |   :|  ||    | :||  |:  :
Db        398 EKQRAEHIM-----------------LVDLG---RNDVGKVSKPGSVNVEKLMSVERYS 436

Qy        514 -----PGTAKSQFLKYIEKVSS-RAIFTTG--QGASAVAVTAYVQRHPVSREWTLEAG-- 563
                    |   : | ::    : ||    |   ||  |     : :  |:|      |
Db        437 HVMHISSTVSGELLDHLTCWDALRAALPVGTVSGAPKVKAMELIDQLEVARRGPYSGGFG 496

Qy        564 --------ALVLADRGVCLID--EFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARC 613
                     : || | :  ::   :| |      |  :| ::|          |  ||:
Db        497 GISFSGDMDIALALRTMVFLNGARYDTM-----YSYTDASKRQE-------WVAHLQSGA 544

Qy        614 TVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHV 673
               ::| :||            | ::    :      ||   |  :       : || |
Db        545 GIVADSNP----------DEEQIECENKVAG----LCRAIDLAE-------SAFVKG--- 580

Qy        674 RHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLN 725
              || || |      |||     :|| :        | ::  : :|:||| | |
Db        581 RHKPSVK-----INGS-----VPNLF------SRVQRQTSVMSKDRVHEKRN 616


RESULT 12
US-09-264-854-5
; Sequence 5, Application US/09264854
; Patent No. 6563025
; GENERAL INFORMATION:
;  APPLICANT: SONG, HEE-SOOK
;  APPLICANT: BROTHERTON, JEFFREY E.
;  APPLICANT: WIDHOLM, JACK M.
;  TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
;  TITLE OF INVENTION: TRANSFORMATION
```

```
; FILE REFERENCE: UI001.C2
; CURRENT APPLICATION NUMBER: US/09/264,854
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: US 09/001,826
; EARLIER FILING DATE: 1997-12-31
; EARLIER APPLICATION NUMBER: US 08/937,739
; EARLIER FILING DATE: 1997-07-25
; EARLIER APPLICATION NUMBER: US 60/025,140
; EARLIER FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text
; SEQ ID NO 5
;   LENGTH: 616
;   TYPE: PRT
;   ORGANISM: Nicotiana tabacum
US-09-264-854-5

  Query Match             2.5%;  Score 117;  DB 4;  Length 616;
  Best Local Similarity   21.0%;  Pred. No. 0.043;
  Matches  137;  Conservative   84;  Mismatches  231;  Indels  200;  Gaps   37;

Qy        148 EDGEEDEQMIESIENLEDLKGHSV-REWVSMAGPRLEI-----------HHRFKNFLRTH 195
Db         91 DDREAPSFLFESVE--PGFRGSSVGRYSVVGAQPSMEIVAKEHNVTILDHHTGKLTQKTV 148

Qy        196 VD-------SHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLP--EAPAELL 245
Db        149 QDPMTIPRSISEG---WKPRLIDELPDTFCGGWVGYFSYDTVRYVENRKLPFLRAPED-- 203

Qy        246 QIFDEAALEVVLAMYPK---YDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVT 302
Db        204 ---DRNLADIQLGLYEDVIVFDHVEKKAHV--IHWVQLDQYSSLPEAYLDGKKRLEILVS 258

Qy        303 SCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSA-GP--FEVNMEETIY 359
Db        259 RVQGI-----------------------ESPRLSPGSVDFCTHAFGPSLTKGNMTSEEY 294

Qy        360 QNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTAN 419
Db        295 KN----AVLQAKEHIAAGDIFQ----IVLSQRFERRTFADPFE---VYR-----ALRIVN 338

Qy        420 GFPVFATVILAN---------HVAKKDNKVAV-----------GELTDEDVKMITSLSKD 459
Db        339 PSP-YMTYIQARGCILVASSPEILTRVKKRRIVNRPLAGTSRRGKTPDEDVMLEMQMLKD 397

Qy        460 Q-QIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINV--LLCGD--- 513
Db        398 EKQRAEHIM------------------LVDLG---RNDVGKVSKPGSVNVEKLMSVERYS 436

Qy        514 -----PGTAKSQFLKYIEKVSS-RAIFTTG--QGASAVAVTAYVQRHPVSREWTLEAG-- 563
Db        437 HVMHISSTVSGELLDHLTCWDALRAALPVGTVSGAPKVKAMELIDQLEVARRGPYSGGFG 496

Qy        564 --------ALVLADRGVCLID--EFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARC 613
Db        497 GISFSGDMDIALALRTMVFLNGARYDTM-----YSYTDASKRQE-------WVAHLQSGA 544

Qy        614 TVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHV 673
Db        545 GIVADSNP----------DEEQIECENKVAG----LCRAIDLAE-------SAFVKG--- 580

Qy        674 RHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLN 725
Db        581 RHKPSVK-----INGS-----VPNLF------SRVQRQTSVMSKDRVHEKRN 616

RESULT 13
US-09-252-991A-17243
; Sequence 17243, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 17243
;   LENGTH: 771
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17243

  Query Match             2.5%;  Score 116;  DB 4;  Length 771;
  Best Local Similarity   19.0%;  Pred. No. 0.08;
  Matches  140;  Conservative   73;  Mismatches  205;  Indels  320;  Gaps   37;

Qy        225 DLAAR---EHVLAYFLPEAPA--------------ELLQIFDEAALEVVLAMYPKYDRI 266
Db         63 DVQARAFVEDVLAHRLARPLAVVLADRTEDGVVFLEDLLGVFRLAAHQVVAAFQGAADLF 122

Qy        267 TNHIHVRISHLPLVEELRSLRQLH---------LNQLIRTSGVVTSCTGVLPQLSMVKYN 317
Db        123 GN-----------AEEDRVRGELHQAPVEGRVGLNQSLRVA------------------ 152

Qy        318 CNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAAR 377
Db        153 ------ILDRFAQVQ----------------------------------QQQPQFADLR 171

Qy        378 RLPRSKDAILLADLVDSCNAGDE-----IELTGIYHNNYDGSLNTANGFPVFATVILANH 432
Db        172 RPGALRD-----------DADDERLQRIAQLQDVFHGRLAGRRDLAQRFAAF-------- 212

Qy        433 VAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRG-PA------ 485
Db        213 ---EDHHALARQRTDH---------------------AQGLQDHDRLAQARPADFHLFS 247

Qy        486 -LALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVA 544
Db        248 QLAL--------AGQH---------LAGDPAT----LLDGVEQALGNRVDQLGRRKHVAS 286

Qy        545 VTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHE-AMEQQSISISK- 602
Db        287 L--YL---PIGRATRLGASAMQLAFPG-----------------HRAALERLAVSLGRR 323

Qy        603 --AGIVTSLQAR--CTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDP 658
Db        324 LQAPVEYGLRARHIDVVFPAGDGHGGHAVADQVGQRATDAHEPV--------------- 367

Qy        659 VQDEMLARFVVGSHVRH-HPS---NKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYII 714
Db        368 ------------HRKHQHQAGHRDRRHRGEGGGEDHQGRAGNAVGA-----------L 402

Qy        715 YAKERVHPKLNQMDQDKVA---KMYSDLRKESMATGSI---PITVR------------- 754
Db        403 GGHQR-----NQQDGDQVGHRQRRAGDVGDEHNRQGQVDRESVEVERVAGGNHQAYRGAA 457

Qy        755 --HIESMSH-------GGGPRAHPSAGLCD-RRRRQHGHP---RDAGELHRHTEVQ---- 797
Db        458 DAHVLELAHHLRQHRVGGGGAHHDQQLFAEVAQERQHAHPGETQDAAE-HRHDEDQAGGV 516

Qy        798 RHRSMRKTFARYL--------SFRRDN-------------NELLLFI-----LKQLVAE 830
Db        517 EHQHQRAQVAERVDAVAGGGEGHRAEHAERCQAHDHAHDAEHDVAEFVDQAGHAAGLLAD 576

Qy        831 QVTYQRNRFGAQQDTIEV 848
Db        577 QVQGAAEQHGEQQDLEDV 594

RESULT 14
```

```
US-08-840-006-6
; Sequence 6, Application US/08840006
; Patent No. 6127520
;  GENERAL INFORMATION:
;    APPLICANT:  Ueda, Tetsufumi
;    APPLICANT:  Ozkan, Eric D.
;    TITLE OF INVENTION:  Compositions And Methods For The
;    TITLE OF INVENTION:  Inhibition Of Neurotransmitter Uptake Of Synaptic Vesicles
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/840,006
;      FILING DATE:  15-APR-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  UM-02786
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1151 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-840-006-6

  Query Match             2.5%;  Score 116;  DB 3;  Length 1151;
  Best Local Similarity   18.2%;  Pred. No. 0.16;
  Matches  179;  Conservative  156;  Mismatches  332;  Indels  318;  Gaps   49;

Qy         35 RDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVE 94
              |||  :  | : |:  :  |  :  | | :   ::|      |:||:| :     ||  :
Db        346 RDLTSWVTEMKALINAD-ELASDVAGAEAL---LDRHQEHKGEIDAHE-DSFKSADESGQ 400

Qy         95 ELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDE 154
               | |:   |:|                |   :| :   |:||          |  |
Db        401 ALLAAGHYASD---------------EVREKLTVLSEERAAL---------LELWELRR 435

Qy        155 QMIESIENLE----DLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISD 210
              |  |   :|:    | :   |  |:|          : : ||             | : |
Db        436 QQYEQCMDLQLFYRDTE--QVDNWMS----------KQEAFL-----------LNEDLGD 472

Qy        211 MCKENRESLVVNYED----LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRI 266
                 :: |:|:  :||    |:|:|             | :   || | :::   :   : :
Db        473 FL-DSVEALLKKHEDFEKSLSAQE-----------EKITALDEFATKLIQNNHYAMEDV 519

Qy        267 TNHIHVRISHLPLVEELRSLRQ-------LHLNQLIRTSGVVTSCTGVLPQLSMVKYNCN 319
                    :|    : | |::|:        || |  | |  : |
Db        520 ATRRDALLSRRNALHE-RAMRRRAQLADSFHLQQFFRDSDELKSWV-------------- 564

Qy        320 KCNFVLGPFCQSQNQEVK---------PGSCP-ECQSAGPFEVNMEETIYQNYQRIRIQE 369
                        |:::|         | :   : |    ||  :     |  ||   |
Db        565 ------------NEKMKTATDEAYKDPSNLQGKVQKHQAFEAELS----ANQSRIDALE 607

Qy        370 SPG-----------KVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNT 417
               |           :||||    ::   |   |:::       || ||         |
Db        608 KAGQKLIDVNHYAKDEVAAR---MNEVISLWKKLLEA------TELKGI-------KLRE 651

Qy        418 ANGFPVFATVI---------LANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFA 468
              ||    |   :          :  |:|  |          :|:  : :| |   : |   |
Db        652 ANQQQQFNRNVEDIELWLYEVEGHLASDD--------YGKDLTNVQNLQKKHALLEADVA 703

Qy        469 SIAPSI-----------YGHEDI----KRGPALALFGGEPKNP--GGKHKVRGDINVLL 510
              :    |              || |     |:  ||       | |    | |:   : :
Db        704 AHQDRIDGVTIQARQFQDAGHFDAENIKKKQEALVARYEALKEPMVARKQKLADSLRLQQ 763

Qy        511 CGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHP------VSREWTLEAGA 564
                     :  :::  | :::    :| :|   : |   :::|          |  ::|
Db        764 LFRDVEDEETWIREKEPIAA----STNRGKDLIGVQNLLKKHQALQAEIAGHEPRIKA-- 817

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAG-----IVTSLQARCTVIAAA 619
                :  :|  :::|     :  :  :||  ::     :||      :  ||||:    | |
Db        818 --VTQKGNAMVEEGHFAAEDVKAKLHELNQKWEALKAKASQRRQDLEDSLQAQ-QYFADA 874

Qy        620 N----------PIGGRYDPSLTFSENVDLTEPIISRFDIL------------------- 649
              |          || |  |    : :: |  | :: : : |
Db        875 NEAESWMREKEPIVGSTD----YGKDEDSAEALLKKHEALMSDLSAYGSSIQALREQAQS 930

Qy        650 CVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE---EEG----LANG----------SAAE 692
              |  |  | |  ||     |:  :     | :|   ::|    | |           :  :
Db        931 C--RQQVAPTDDETGKELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQ 988

Qy        693 PAMPNTYGVEPL--PQEVLKKYIIYAKERVHPKLNQMD-QDKVAKMYSDLRKESMATGSI 749
                :|  | |: |   |   :: ::  :  :  :  |:| | :: |           ||:
Db        989 GFVPAAY-VKKLDPAQSASRENLLEEQGSIALRQEQIDNQTRITK----------EAGSV 1037

Qy        750 PITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARY 809
              : :: :| : |                          : ||       :|   : |: ::
Db       1038 SLRMKQVEELYHS--------------------LLELGE-------KRKGMLEKSCKKF 1069

Qy        810 LSFRRDNNELLLFILKQ---LVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLS 866
              : | |: |||  :| ::   | :|:|       ||  : :|| :|   |  : :  :
Db       1070 MLF-REANELQQWINEKEAALTSEEV-------GADLEQVEVLQKKFDDFQKDLKAN--- 1118

Qy        867 AFYDSELFRMNKFSHDLKRKMILQQ 891
                :| |  :|| : ||: : :: :
Db       1119 ---ESRLKDINKVAEDLESEGLMAE 1140


RESULT 15
US-08-840-006-5
; Sequence 5, Application US/08840006
; Patent No. 6127520
;  GENERAL INFORMATION:
;    APPLICANT:  Ueda, Tetsufumi
;    APPLICANT:  Ozkan, Eric D.
;    TITLE OF INVENTION:  Compositions And Methods For The
;    TITLE OF INVENTION:  Inhibition Of Neurotransmitter Uptake Of Synaptic Vesicles
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/840,006
;      FILING DATE:  15-APR-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  UM-02786
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
```

```
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1200 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-840-006-5

  Query Match             2.5%;  Score 116;  DB 3;  Length 1200;
  Best Local Similarity   18.2%;  Pred. No. 0.18;
  Matches  179;  Conservative  156;  Mismatches  332;  Indels  318;  Gaps   49;

Qy         35 RDLPPFEDESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVE  94
              |||  :  | : |:  :  |  :  | | :   ::|      |:||:| :     ||  :
Db        371 RDLTSWVTEMKALINAD-ELASDVAGAEAL---LDRHQEHKGEIDAHE-DSFKSADESGQ 425

Qy         95 ELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDE 154
               | |:   |:|                  |   :| :   |:||          |  |
Db        426 ALLAAGHYASD---------------EVREKLTVLSEERAAL---------LELWELRR 460

Qy        155 QMIESIENLE----DLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISD 210
              |  |   :|:    | :   |  |:|          : : ||             | : |
Db        461 QQYEQCMDLQLFYRDTE--QVDNWMS----------KQEAFL-----------LNEDLGD 497

Qy        211 MCKENRESLVVNYED----LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRI 266
                :: |:|:  :||    |:|:|           | :   || | :::   :   : :
Db        498 FL-DSVEALLKKHEDFEKSLSAQE-----------EKITALDEFATKLIQNNHYAMEDV 544

Qy        267 TNHIHVRISHLPLVEELRSLRQ-------LHLNQLIRTSGVVTSCTGVLPQLSMVKYNCN 319
                    :|    : | |::|:         || |  | |  : |
Db        545 ATRRDALLSRRNALHE-RAMRRRAQLADSFHLQQFFRDSDELKSWV-------------- 589

Qy        320 KCNFVLGPFCQSQNQEVK---------PGSCP-ECQSAGPFEVNMEETIYQNYQRIRIQE 369
                        |:::|          | :   : |    ||  :       |  ||   |
Db        590 -------------NEKMKTATDEAYKDPSNLQGKVQKHQAFEAELS----ANQSRIDALE 632

Qy        370 SPG-----------KVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNT 417
               |           :||||    ::   |   |:::       || ||          |
Db        633 KAGQKLIDVNHYAKDEVAAR---MNEVISLWKKLLEA------TELKGI-------KLRE 676

Qy        418 ANGFPVFATVI---------LANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFA 468
              ||   |  :          :  |:|  |          :|:  : :| |   : |   |
Db        677 ANQQQQFNRNVEDIELWLYEVEGHLASDD--------YGKDLTNVQNLQKKHALLEADVA 728

Qy        469 SIAPSI-----------YGHEDI----KRGPALALFGGEPKNP--GGKHKVRGDINVLL 510
              :    |           || |     |:  ||       | |    | |:   : :
Db        729 AHQDRIDGVTIQARQFQDAGHFDAENIKKKQEALVARYEALKEPMVARKQKLADSLRLQQ 788

Qy        511 CGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHP------VSREWTLEAGA 564
                 :  :::  | :::    :| :|   : |   :::|          |  ::|
Db        789 LFRDVEDEETWIREKEPIAA----STNRGKDLIGVQNLLKKHQALQAEIAGHEPRIKA-- 842

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAG-----IVTSLQARCTVIAAA 619
              :  :|  :::|    :  :  :|| ::     :||       :  ||||:    | |
Db        843 --VTQKGNAMVEEGHFAAEDVKAKLHELNQKWEALKAKASQRRQDLEDSLQAQ-QYFADA 899

Qy        620 N----------PIGGRYDPSLTFSENVDLTEPIISRFDIL------------------- 649
              |          || |  |    : :: |  | :: : : |
Db        900 NEAESWMREKEPIVGSTD----YGKDEDSAEALLKKHEALMSDLSAYGSSIQALREQAQS 955

Qy        650 CVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE---EEG----LANG----------SAAE 692
              |  |  | |  ||     |:  :     | :|   ::|    | |             :  :
Db        956 C--RQQVAPTDDETGKELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQ 1013

Qy        693 PAMPNTYGVEPL--PQEVLKKYIIYAKERVHPKLNQMD-QDKVAKMYSDLRKESMATGSI 749
              :|  | |: |   |   :: ::  :  :  :  |:| | :: |           ||:
Db       1014 GFVPAAY-VKKLDPAQSASRENLLEEQGSIALRQEQIDNQTRITK----------EAGSV 1062

Qy        750 PITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARY 809
               : :: :| : |                        : ||       :|   : |:  ::
Db       1063 SLRMKQVEELYHS--------------------LLELGE-------KRKGMLEKSCKKF 1094

Qy        810 LSFRRDNNELLLFILKQ---LVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLS 866
              : | |: |||  :| ::   | :|:|       ||  : :|| :|   |  : :  :
Db       1095 MLF-REANELQQWINEKEAALTSEEV-------GADLEQVEVLQKKFDDFQKDLKAN--- 1143

Qy        867 AFYDSELFRMNKFSHDLKRKMILQQ 891
                :| |  :|| : ||: : :: :
Db       1144 ---ESRLKDINKVAEDLESEGLMAE 1165


Search completed: September 12, 2003, 13:21:29
Job time : 26 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2003, 13:13:55 ; Search time 70 Seconds
(without alignments)
1859.341 Million cell updates/sec

Title: CAA47749
Perfect score: 4630
Sequence: 1 MASSPAQRRRGNDPLTSSPG.....LFRMNKFSHDLKRKMILQQF 892

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4078.5	88.1	904	10	US-09-964-899-49	Sequence 49, Appl
2	1872	40.4	864	15	US-10-128-714-8071	Sequence 8071, Ap
3	1840.5	39.8	785	15	US-10-128-714-3071	Sequence 3071, Ap
4	1073.5	23.2	268	10	US-09-925-300-1461	Sequence 1461, Ap
5	891	19.2	1023	15	US-10-128-714-8558	Sequence 8558, Ap
6	887	19.2	903	15	US-10-128-714-8363	Sequence 8363, Ap
7	882.5	19.1	831	15	US-10-128-714-3363	Sequence 3363, Ap
8	860	18.6	874	15	US-10-128-714-3558	Sequence 3558, Ap
9	835	18.0	821	15	US-10-171-311-129	Sequence 129, App
10	835	18.0	901	11	US-09-828-062-8	Sequence 8, Appli
11	810	17.5	718	15	US-10-128-714-3399	Sequence 3399, Ap
12	804	17.4	713	15	US-10-128-714-8399	Sequence 8399, Ap
13	783	16.9	724	10	US-09-925-300-1053	Sequence 1053, Ap
14	778	16.8	766	9	US-09-925-301-1382	Sequence 1382, Ap
15	776	16.8	921	15	US-10-128-714-3372	Sequence 3372, Ap
16	774	16.7	720	15	US-10-101-080-2	Sequence 2, Appli
17	763	16.5	819	9	US-09-833-790-425	Sequence 425, App
18	754	16.3	957	15	US-10-128-714-8372	Sequence 8372, Ap
19	753.5	16.3	814	15	US-10-128-714-8100	Sequence 8100, Ap
20	732.5	15.8	632	15	US-10-128-714-3100	Sequence 3100, Ap
21	451	9.7	94	9	US-09-925-299-1224	Sequence 1224, Ap
22	451	9.7	94	11	US-09-925-299-1224	Sequence 1224, Ap
23	451	9.7	113	15	US-10-106-698-5543	Sequence 5543, Ap
24	236	5.1	69	9	US-09-864-761-46505	Sequence 46505, A
25	201	4.3	171	15	US-10-106-698-5506	Sequence 5506, Ap
26	193.5	4.2	63	9	US-09-864-761-45117	Sequence 45117, A
27	186	4.0	66	9	US-09-925-302-730	Sequence 730, App
28	154	3.3	138	10	US-09-764-868-939	Sequence 939, App
29	139	3.0	178	15	US-10-106-698-6566	Sequence 6566, Ap
30	132.5	2.9	367	10	US-09-738-626-5688	Sequence 5688, Ap
31	129.5	2.8	872	14	US-10-047-260-38	Sequence 38, Appl
32	127	2.7	507	10	US-09-738-626-5725	Sequence 5725, Ap
33	124	2.7	867	14	US-10-007-693-109	Sequence 109, App
34	123	2.7	1177	15	US-10-128-714-3493	Sequence 3493, Ap
35	121	2.6	1179	15	US-10-128-714-8493	Sequence 8493, Ap
36	120.5	2.6	856	9	US-09-815-242-11310	Sequence 11310, A
37	120.5	2.6	1324	12	US-10-205-219-113	Sequence 113, App
38	119.5	2.6	541	15	US-10-156-761-10169	Sequence 10169, A
39	117.5	2.5	856	9	US-09-815-242-11489	Sequence 11489, A
40	117	2.5	616	15	US-10-138-927-103	Sequence 103, App
41	115	2.5	1481	15	US-10-331-061-70	Sequence 70, Appl
42	115	2.5	2910	14	US-10-124-800-2	Sequence 2, Appli
43	113.5	2.5	882	9	US-09-815-242-10668	Sequence 10668, A
44	113.5	2.5	6304	15	US-10-147-026-16	Sequence 16, Appl
45	113	2.4	2472	9	US-09-815-242-5064	Sequence 5064, Ap

ALIGNMENTS

RESULT 1
US-09-964-899-49
; Sequence 49, Application US/09964899
; Patent No. US20020174446A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/298,309
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-964-899-49

Query Match 88.1%; Score 4078.5; DB 10; Length 904;
Best Local Similarity 89.8%; Pred. No. 0;
Matches 817; Conservative 8; Mismatches 52; Indels 33; Gaps 6;

```
Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 69

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAA-------DGPCGTVTG 113
              ||||||||||||||||||||||||||||||||||||||||:||||       |   |
Db         70 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRQRDREAGR--- 126

Qy        114 SWPGLGACAVGSCMTA--------MRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
                 |||    |    :         |: |    |       ||||||||:||||||||||
Db        127 ---GLGRMRRGLLYDSDEEDEERPARKRRQVERA-------TEDGEEDEEMIESIENLED 176
```

```
Qy        166 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 LKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYED 236

Qy        226 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 LAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRS 296

Qy        286 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 LRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQ 356

Qy        346 SAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTG 405
              |||||||||||||||||||||||||||| |||||||||||||||  ||||||||
Db        357 SAGPFEVNMEETIYQNYQRIRIQESPGKVAAGRLPRSKDAILLADLVDSCKPGDEIELTG 416

Qy        406 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 IYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEK 476

Qy        466 IFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 525
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        477 IFASIAPSIYGHEDIKRGLALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYI 536

Qy        526 EKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 585
              |||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||||
Db        537 EKVSSRAIFTTGQGASAVGLTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQD 596

Qy        586 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 RTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISR 656

Qy        646 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        657 FDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLP 716

Qy        706 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMSHGGGP 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 QEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESMIRMA-- 774

Qy        766 RAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRH---RSMRKTFARYLSFRRDNNELLLF 822
              ||    | |       :      |    : |:    ||||||||||||||||||||||||
Db        775 EAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFSVMRSMRKTFARYLSFRRDNNELLLF 834

Qy        823 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 ILKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQINIHNLSAFYDSELFRMNKFSHD 894

Qy        883 LKRKMILQQF 892
              ||||||||||
Db        895 LKRKMILQQF 904
```

```
RESULT 2
US-10-128-714-8071
; Sequence 8071, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8071
;   LENGTH: 864
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-8071

  Query Match             40.4%;  Score 1872;  DB 15;  Length 864;
  Best Local Similarity   48.0%;  Pred. No. 4.4e-155;
  Matches  399;  Conservative  110;  Mismatches  213;  Indels  110;  Gaps   12;

Qy         14 PLTSSP----GRSSRR-------TDALTSSPGRDLPPF-------------EDESEGLLG 49
              | ||:     || :||        :| : :|    ||                ||   |||
Db          7 PSTSAANRGLGRLNRRKRSREPDDEASSVAPPSSAPPSSPPMLPLDEDDIERDEEAELLG 66

Qy         50 TEGPLE---EEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADG 106
                ::   |:||| :| ||  ||||| :   | || | :  || | |||  : |   :
Db         67 DIDDIDEMAEDEDGIDLFGDNFERDYRDVGN-DRYEGEYID-DDGDHEELDIATRRQLEA 124

Qy        107 PCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEE-------------- 152
                                    || |           |  :| :|
Db        125 RLN--------------------RRDRELERRRRMPAAFLQDDDEGDMDLTRQPRRRRH 163

Qy        153 ------------DEQMIE-SIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSH 199
                         |: | | |:| | |:|  :: :||        |:  || ||    |
Db        164 HYDEDREDIEMGDDAMEELSLEELADVKAANITDWVLQPQVLRSIYREFKAFLTEFTDQA 223

Qy        200 GHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAM 259
              | :|:  :|  : : |  || |:|  |:: :  |:|||   | |:|::||: ||:| |
Db        224 GQSVYGHKIKTLGEVNSASLEVSYVHLSSTKAALSYFLANEPTEVLKVFDQVALDVTLFH 283

Qy        260 YPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCN 319
              ||:|  |   ||||||| ||:|  || ||| ||| |:| |||||  ||| |||  | : |
Db        284 YPQYQDIHKEIHVRISDLPIVYTLRQLRQQHLNCLVRVSGVVTRRTGVFPQLKYVMFICQ 343

Qy        320 KCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRL 379
              |||  |||| |  : |||   |  ||| ||| :| |:|:|:|||:: :||||| | | ||
Db        344 KCNITLGPFQQEASAEVKISYCQNCQSKGPFTINSEKTVYRNYQKLTLQESPGSVPAGRL 403

Qy        380 PRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK 439
              || :: :|||||:||   |||||:|||| |:||  ||  |||||||:| |||: |  ::
Db        404 PRQREVVLLADLIDSAKPGDEIEVTGIYRNSYDAQLNNKNGFPVFATIIEANHIVKSHDQ 463

Qy        440 VAVGELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGK 499
              :|   ||:|| : | :||:|  | :||  |||||||||:|:|   ||:|||| |   ||
Db        464 LAGFHLTEEDERQIRALSRDPDIVDKIVRSIAPSIYGHQDVKTAVALSLFGGVSKEAQGK 523

Qy        500 HKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWT 559
                :|||||||| |||||||| |||:|| : ||:| |||||||| :|| |:| |:: |||  |
Db        524 MAIRGDINVLLLGDPGTAKSQVLKYVEKTAHRAVFATGQGASAVGLTASVRRDPLTSEWT 583

Qy        560 LEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAA 619
              || |||||||| ||||||||||||||||||||||:|||||||||:||||| |:|||  |
Db        584 LEGGALVLADRGTCLIDEFDKMNDQDRTSIHEAMEQQTISISKAGIVTTLQARCAVVAAA 643

Qy        620 NPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSN 679
              |||||||: :: ||:||:|||||:|||||||||||||| :|| || ||: || | :|:
Db        644 NPIGGRYNSTIPFSQNVELTEPILSRFDILCVVRDTVDPNEDERLANFVIESHHRANPTR 703

Qy        680 K----------------EEEG----------------LANGSAAEPAMPNTYGVEPLPQE 707
                              :|||                :|   ||:         | :|||
Db        704 PLRDQDGNLVDSEGNRIDEEGYRLDKHGNRLPPTPEEIAKREAAQRKAEEEKEGE-IPQE 762

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATGSIPITVRHIESM 759
              :|:|||:||:|: |||| |:|||||::::|:|:||:|||: |||||||:|::
```

```
Db        763 LLRKYILYAREKCHPKLYQIDQDKVARLFADMRRESLATGAYPITVRHLEAI 814


RESULT 3
US-10-128-714-3071
; Sequence 3071, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3071
;   LENGTH: 785
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-3071

  Query Match            39.8%;  Score 1840.5;  DB 15;  Length 785;
  Best Local Similarity  49.0%;  Pred. No. 2.2e-152;
  Matches  388;  Conservative  102;  Mismatches  207;  Indels   95;  Gaps   10;

Qy         33 PGRDLPPF---------EDESEGLLGTEGPLE---EEEDGEELIGDGMERDYRAIPELDA 80
              |    ||           ||   |||    ::   |:||| :| ||  ||||| :   |
Db          5 PPPSSPPMLPLDEDDIERDEEAELLGDIDDIDEMAEDEDGIDLFGDNFERDYRDVGN-DR 63

Qy         81 YEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPAS 140
              || | :  || | |||  : |   :                              || |
Db         64 YEGEYID-DDGDHEELDIATRRQLEARLN--------------------RRDRELERRR 101

Qy        141 AASGAGTEDGEE-------------------------DEQMIE-SIENLEDLKGHSVRE 173
                 |  :| :|                         |: | | |:| | |:|  :: :
Db        102 RMPAAFLQDDDEGDMDLTRQPRRRRHHYDEDREDIEMGDDAMEELSLEELADVKAANITD 161

Qy        174 WVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVL 233
              ||         |:  || ||    |  | :|:  :|  : : |  || |:|  |:: :  |
Db        162 WVLQPQVLRSIYREFKAFLTEFTDQAGQSVYGHKIKTLGEVNSASLEVSYVHLSSTKAAL 221

Qy        234 AYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQ 293
              :|||   | |:|::|: ||:| |  ||:|  |   |||||| ||:|  || ||| |||
Db        222 SYFLANEPTEVLKVFDQVALDVTLFHYPQYQDIHKEIHVRISDLPIVYTLRQLRQQHLNC 281

Qy        294 LIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPECQSAGPFEVN 353
              |:| |||||  ||| |||  | : | |||  |||| |  : |||   |  ||| ||| :|
Db        282 LVRVSGVVTRRTGVFPQLKYVMFICQKCNITLGPFQQEASAEVKISYCQNCQSKGPFTIN 341

Qy        354 MEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDG 413
               |:|:|:|||:: :|||||  | | |||| :: :|||||:||   |||||:|||| |:||
Db        342 SEKTVYRNYQKLTLQESPGSVPAGRLPRQREVVLLADLIDSAKPGDEIEVTGIYRNSYDA 401

Qy        414 SLNTANGFPVFATVILANHVAKKDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASIAPS 473
               ||  ||||||||:| |||: |  :::|   ||:|| : | :||:|  | :||  ||||
Db        402 QLNNKNGFPVFATIIEANHIVKSHDQLAGFHLTEEDERQIRALSRDPDIVDKIVRSIAPS 461

Qy        474 IYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAI 533
              ||||:|:|   ||:|||| |   ||  :|||||||| |||||||| |||:|| : ||:
Db        462 IYGHQDVKTAVALSLFGGVSKEAQGKMAIRGDINVLLLGDPGTAKSQVLKYVEKTAHRAV 521

Qy        534 FTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAM 593
              | |||||||| :|| |:| |:: |||| ||||||||| ||||||||||||||||||||
Db        522 FATGQGASAVGLTASVRRDPLTSEWTLEGGALVLADRGTCLIDEFDKMNDQDRTSIHEAM 581

Qy        594 EQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVR 653
              |||:||||||||||:|||| |:|||||||||: :: ||:||:|||||:||||||||||
Db        582 EQQTISISKAGIVTTLQARCAVVAAANPIGGRYNSTIPFSQNVELTEPILSRFDILCVVR 641

Qy        654 DTVDPVQDEMLARFVVGSHVRHHPSNK----------------EEEG------------- 684
              ||||| :|| || ||: || | :|:                   :|||
Db        642 DTVDPNEDERLANFVIESHHRANPTRPLRDQDGNLVDSEGNRIDEEGYRLDKHGNRLPPT 701

Qy        685 ---LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK 741
                 :|   ||:          | :|||:|:|||:||:|: |||| |:||||||::::|:|:
Db        702 PEEIAKREAAQRKAEEEKEGE-IPQELLRKYILYAREKCHPKLYQIDQDKVARLFADMRR 760

Qy        742 ESMATGSIPITV 753
              ||:|||: ||||
Db        761 ESLATGAYPITV 772


RESULT 4
US-09-925-300-1461
; Sequence 1461, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1461
;   LENGTH: 268
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-300-1461

  Query Match            23.2%;  Score 1073.5;  DB 10;  Length 268;
  Best Local Similarity  91.7%;  Pred. No. 9e-86;
  Matches  211;  Conservative    5;  Mismatches   13;  Indels    1;  Gaps    1;

Qy          1 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         30 MASSPAQRRRGNDPLTSSPGRSSRRTDALTSSPGRDLPPFEDESEGLLGTEGPLEEEEDG 89

Qy         61 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGTVTGSWPGL-G 119
              |||||||||||||||||||||||||||||||||||||||:||||:     |||   |
Db         90 EELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASQREAAERAMRHVTGRLAGAWA 149

Qy        120 ACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAG 179
              |||||||||||||||||||||||||||| ||||||||:|||||||||||||||||||||
Db        150 ACAVGSCMTAMRRTRSALPASAASGAATEDGEEDEEMIESIENLEDLKGHSVREWVSMAG 209

Qy        180 PRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDLAAR 229
              ||||||||||||||||||||||||||||||||||||||||||||||||  ::
Db        210 PRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKENRESLVVNYEDTGSQ 259


RESULT 5
US-10-128-714-8558
; Sequence 8558, Application US/10128714
; Publication No. US20030119013A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8558
;   LENGTH: 1023
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-8558

  Query Match             19.2%;  Score 891;  DB 15;  Length 1023;
  Best Local Similarity   28.7%;  Pred. No. 9.3e-69;
  Matches  268;  Conservative  140;  Mismatches  304;  Indels  222;  Gaps   29;

Qy          1 MASSPAQRRRG---NDPLTSSPGRSSRRTDALTSSPGRDLPPFEDE-SEGLLGTEGPLEE 56
              |:|  : ||||    |  |||| ||:|     ||||    |   || |:        |
Db          1 MSSPASSRRRGRPAKDTATSSPARSTRSRQLQTSSP---TPRAADEQSQATPRASRRLR- 56

Qy         57 EEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAAD---------GP 107
                 ||  :       ::: |      ||     |  ::| ::  ||::          |
Db         57 ---GEAAVPSSSPMFFQSSPSKADSSAE---TPDVRMDEPSSPMRESSTMDEGDRTPRGN 110

Qy        108 CGTVTGSWPGLGACAVGSCMTAMRR--TRSALPASA---------------------- 141
              |:  | | :   :  |   |  |   || :|:|:
Db        111 APTMRDSSP-IRYMSSSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLH 169

Qy        142 -----------------ASGAGTEDGEEDEQMIESIENL-------EDLKGHSVRE-WVS 176
                              |:|    ||  | :   :  |:       | | | | |  |
Db        170 SGGFLSSPNRRRRVFVDANGMPATDG--DPRSDATFSNIHPDTSEAEALGGSSTRVIW-- 225

Qy        177 MAGPRLEIH---HRFKNFL-----RTHVDSHGHNVFKERISDMCKENRE---------- 217
                |  : |      |||||      :  : : |    : ||     | ||
Db        226 --GTNISIQDSMSAFKNFLYNFQTKYRLWAEGATEDETRIMGDSAEEREYISMLSTMRQL 283

Qy        218 ---SLVVNYEDLAAREHVLAYF--LPEAPAELLQIFDEAALEVVLAM------------ 259
                || :: ::| |    |  :  |   | |:: : |:   :|:: :
Db        284 GVTSLNLDAKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQ 343

Qy        260 --------------YPKYDRITNHIHVRISHLP-LVEE-----------------LRSL 286
                           |  | ::    :  : :| || |                   :| |
Db        344 RNQNHNRGLSSGPAVPSSDALSETGRMPQNEIPDLVGEVETKAFKVLPFGLDSTVNMRDL 403

Qy        287 RQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSCPE--C 344
                :::|:   |:|   | ::| :    : |  ||  :       : :  :|  ||   |
Db        404 DPADMDKLVSIKGLVIRTTPIIPDMKEAFFRCQVCNHGV-QVDIDRGKIAEPTECPRPVC 462

Qy        345 QSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELT 404
              :    ::     :: : | |::|:|  :   : | |     :  :||| | ||| :|:|
Db        463 KERNSMQLIHNRCVFADKQVIKLQETPDSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVT 522

Qy        405 GIYHNNYDGSLNTANGFPV-----------FATVILANHVAKKDNK------------ 439
              ||:  |           ||            | | |   || | | |
Db        523 GIFRCN-----------PVRVNPRQRTQKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQEL 571

Qy        440 --VAVGE------LTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGG 491
                  | |:      || |: : |   :    : | :  |:|||||  :|:|:|  | ||||
Db        572 SEQAAGDAEQTRRLTAEEEEKIKRTATRPDLYELLSRSLAPSIYEMDDVKKGILLQLFGG 631

Qy        492 EPK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYV 549
                 |    ||  : |||||:|||||| |:||| |:|: |:: | ::|:|:|:|||| :|||| |
Db        632 TNKTFQKGGNPRYRGDINILLCGDPSTSKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYV 691

Qy        550 QRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSL 609
                | | :|:  ||:|||||:| |:| |||||||:  |: :|| ||||::||:||||:|:|
Db        692 TRDPETRQMVLESGALVLSDGGICCIDEFDKMNESTRSVLHEVMEQQTVSIAKAGIITTL 751

Qy        610 QARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVV 669
                || :::|:|||| ||:|:|   :|:||   ::|||:: :| | ||  :|  ||: :|
Db        752 NARTSILASANPIGSRYNPNLPVPQNIDLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLV 811

Qy        670 GSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQ 729
                ::   | :  |:                  | || | |  || ||| :||| |
Db        812 NMYLEDRPEHAAEQ-----------------EILPIEFLTAYITYAKTKVHPVLTPAAG 853

Qy        730 DKVAKMYSDLRKES----MATGSIPITVRHIESM 759
                ::  | ::||       :   |  | | :|||
Db        854 KALSDAYVNMRKLGDDIRSSDRRITATTRQLESM 887


RESULT 6
US-10-128-714-8363
; Sequence 8363, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8363
;   LENGTH: 903
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-8363

  Query Match             19.2%;  Score 887;  DB 15;  Length 903;
  Best Local Similarity   33.3%;  Pred. No. 1.7e-68;
  Matches  230;  Conservative  109;  Mismatches  252;  Indels  100;  Gaps   13;

Qy        221 VNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRIT-NHIHVRISHLPL 279
                |: ::: |    ||  |  :| :    || |  ||:  :  :  | | : ::   :::
Db         66 VSIDEIRAHNRELADGLLTSPFDYSLAFDRALKEVIKTLPNRPSRETADDVNYYCAYVGA 125

Qy        280 VEEL----RSLRQLHLNQLIRTSGVVTSCTGVLPQ-LSMVKYNCNKCNFVLGPFCQS--- 331
                |      |:|   |||::|   |:|| |: | |: :  | ||  |  ||   :
Db        126 FGEFSCNPRTLGSTHLNRMISLEGIVTKCSLVRPKIIQSVHYNERKDRFVARRYRDQTMT 185

Qy        332 ----QNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAIL 387
```

```
Db        186 TTGITNMNVYP---QEDDEKNPLITEYGYSTYLDHQTISIQEMPERAPAGQLPRSVDVIL 242

Qy        388 LADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK----VAVG 443

Db        243 DDDLVDSAKPGDRIQLVGIYRSL--GNRNASSGSSTFRTVVMANNIIQLSSKSGGGIAQA 300

Qy        444 ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVR 503

Db        301 TITDTDIRNINKVAKKKNVFELLANSLAPSIYGHDYIKKAILLMLLGGMEKNLDNGTHLR 360

Qy        504 GDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAG 563

Db        361 GDINILMVGDPSTAKSQLLRFVLNTAPLAIATTGRGSSGVGLTAAVTSDKETGERRLEAG 420

Qy        564 ALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIG 623

Db        421 AMVLGDRGVVCIDEFDKMSDVDRVAIHEVMEQQTVTIAKAGIHTSLNARCSVLAAANPIY 480

Qy        624 GRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE-- 681

Db        481 GQYDPHKDPHKNIALPDSLLSRFDLLFVVTDDIEDARDRMVSEHVLRMHRYRQPGTEEGA 540

Qy        682 ----------------------------------EEGLANGSAAEPAMPNTYGVEPLPQE 707

Db        541 PVREDLNQTLGVGLEDNQDSNQPTEVYEKFNVMLHAGMANSSR------KGKGIEILSIP 594

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATG---SIPITVRHIESMSHGGG 764

Db        595 FIKKYIQYSKSRIKPVLTKGAADHIVATYSALRNDELSGNQRRTSPITARTLETLIRLS- 653

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFIL 824

Db        654 -TAHAKARLSNRVE----------------ERDAKVAESILRFAMFKE---------- 684

Qy        825 KQLVAEQVTYQRNRFGAQQDTIEVPEKDLVD 855

Db        685 ---VLEDERRKRRKVTTFDEDSESSESDSED 712


RESULT 7
US-10-128-714-3363
; Sequence 3363, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3363
;   LENGTH: 831
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-3363

  Query Match           19.1%;  Score 882.5;  DB 15;  Length 831;
  Best Local Similarity   33.3%;  Pred. No. 3.7e-68;
  Matches  230;  Conservative  109;  Mismatches  251;  Indels  101;  Gaps   14;

Qy        221 VNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRIT-NHIHVRISHLPL 279

Db         28 VSIDEIRAHNRELADLL-TSPFDYSLAFDRALKEVIKTLPNRPSRETADDVNYYCAYVGA 86

Qy        280 VEEL----RSLRQLHLNQLIRTSGVVTSCTGVLPQ-LSMVKYNCNKCNFVLGPFCQS--- 331

Db         87 FGEFSCNPRTLGSTHLNRMISLEGIVTKCSLVRPKIIQSVHYNERKDRFVARRYRDQTMT 146

Qy        332 ----QNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAIL 387

Db        147 TTGITNMNVYP---QEDDEKNPLITEYGYSTYLDHQTISIQEMPERAPAGQLPRSVDVIL 203

Qy        388 LADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNK----VAVG 443

Db        204 DDDLVDSAKPGDRIQLVGIYRSL--GNRNASSGSSTFRTVVMANNIIQLSSKSGGGIAQA 261

Qy        444 ELTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVR 503

Db        262 TITDTDIRNINKVAKKKNVFELLANSLAPSIYGHDYIKKAILLMLLGGMEKNLDNGTHLR 321

Qy        504 GDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAG 563

Db        322 GDINILMVGDPSTAKSQLLRFVLNTAPLAIATTGRGSSGVGLTAAVTSDKETGERRLEAG 381

Qy        564 ALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIG 623

Db        382 AMVLGDRGVVCIDEFDKMSDVDRVAIHEVMEQQTVTIAKAGIHTSLNARCSVLAAANPIY 441

Qy        624 GRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKE-- 681

Db        442 GQYDPHKDPHKNIALPDSLLSRFDLLFVVTDDIEDARDRMVSEHVLRMHRYRQPGTEEGA 501

Qy        682 ----------------------------------EEGLANGSAAEPAMPNTYGVEPLPQE 707

Db        502 PVREDLNQTLGVGLEDNQDSNQPTEVYEKFNVMLHAGMANSSR------KGKGIEILSIP 555

Qy        708 VLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMATG---SIPITVRHIESMSHGGG 764

Db        556 FIKKYIQYSKSRIKPVLTKGAADHIVATYSALRNDELSGNQRRTSPITARTLETLIRLS- 614

Qy        765 PRAHPSAGLCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFARYLSFRRDNNELLLFIL 824

Db        615 -TAHAKARLSNRVE----------------ERDAKVAESILRFAMFKE---------- 645

Qy        825 KQLVAEQVTYQRNRFGAQQDTIEVPEKDLVD 855

Db        646 ---VLEDERRKRRKVTTFDEDSESSESDSED 673


RESULT 8
US-10-128-714-3558
; Sequence 3558, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
```

```
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3558
;   LENGTH: 874
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-3558

  Query Match             18.6%;  Score 860;  DB 15;  Length 874;
  Best Local Similarity   29.6%;  Pred. No. 3.8e-66;
  Matches  251;  Conservative  127;  Mismatches  269;  Indels  200;  Gaps   25;

Qy          2 ASSPA-----QRRRGNDPLTSS-----------PGRS-SRRTD----ALTSSPGRDLPPF 40
              :|||      | || : | :||             |: |||:|       ||| |    |
Db          3 SSSPTRAQNRQSRRSDIPSSSSGLFVSSRPSIESNRAVSRRSDLHSGGFLSSPNRRRRVF 62

Qy         41 EDESEGLLGTEGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASR 100
               | : |:  |:|                       |  ||      :  |  : | |  |
Db         63 VD-ANGMPATDGD---------------------PRSDA-TFSNIHPDTSEAEALGGSS 98

Qy        101 REAADGPCGTVTGSWPGLGACAVGSCMTAMR------RTRSALPASAASGAGTEDGEEDE 154
                  |   ::  |            |:| :      :|:  | |  |    |||  |
Db         99 TRVIWGTNISIQDS------------MSAFKNFLYNFQTKYRLWAEGA----TED--ETR 140

Qy        155 QMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHVDSHGHNVFKERISDMCKE 214
               | :| |          ||::||                                 :| | :
Db        141 IMGDSAEE---------REYISM-----------------------------LSTMRQL 161

Qy        215 NRESLVVNYEDLAAREHVLAYF--LPEAPAELLQIFDEAALEVVLAMYPK--------YD 264
                 || :: ::| |    |  :  |   | |:: : |:   :|:: :  |
Db        162 GVTSLNLDAKNLKAYPSTLKLWHQLHAYPQEIIPLMDQTVKDVMVELAIKEMERLRAQNQ 221

Qy        265 RITNH-----------IHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSM 313
              .  |  ||           |    |     :| |    :::|:   |:|   | ::| :
Db        222 RNQNHNRGLSSVETKAFKVLPFGLDSTVNMRDLDPADMDKLVSIKGLVIRTTPIIPDMKE 281

Qy        314 VKYNCNKCNFVLGPFCQSQNQEVKPGSCPE--CQSAGPFEVNMEETIYQNYQRIRIQESP 371
               : |  ||  :      : :  :|  ||   |:     ::     :: : | |::||:|
Db        282 AFFRCQVCNHGV-QVDIDRGKIAEPTECPRPVCKERNSMQLIHNRCVFADKQVIKLQETP 340

Qy        372 GKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPV-------- 423
               :   | | |    :  :|| | ||| :|:|||:  |           ||
Db        341 DSIPDGQTPHSVSLCVYDELVDVCKAGDRVEVTGIFRCN----------PVRVNPRQRT 389

Qy        424 ----FATVILANHVAKKDNK--------------VAVGE------LTDEDVKMITSLSK 458
                  | | |   || | | |                | |:      || |: : |   :
Db        390 QKSLFKTYIDVLHVQKIDRKKLGIDVSTIEQELSEQAAGDAEQTRRLTAEEEEKIKRTAT 449

Qy        459 DQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPK--NPGGKHKVRGDINVLLCGDPGT 516
                 : | :  |:|||||  :|:|:|  | ||||  |    ||  : |||||:|||||| |
Db        450 RPDLYELLSRSLAPSIYEMDDVKKGILLQLFGGTNKTFQKGGNPRYRGDINILLCGDPST 509

Qy        517 AKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLID 576
              :||| |:|: |:: | ::|:|:|:||| :|||| | | :|:  ||:|||||:| |:| ||
Db        510 SKSQLLRYVHKIAPRGVYTSGKGSSAVGLTAYVTRDPETRQMVLESGALVLSDGGICCID 569

Qy        577 EFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENV 636
              ||||||:  |: :|| ||||::||:||||:|:| || :::|:|||||  ||:|:|   :|:
Db        570 EFDKMNESTRSVLHEVMEQQTVSIAKAGIITTLNARTSILASANPIGSRYNPNLPVPQNI 629

Qy        637 DLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEGLANGSAAEPAMP 696
              ||   ::||||:: :| | ||  :|  ||: :|  ::   | :  |:
Db        630 DLPPTLLSRFDLVYLVLDRVDEQEDRRLAKHLVNMYLEDRPEHAAEQ------------- 676

Qy        697 NTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKES----MATGSIPIT 752
                  | || | |  || ||| :||| |       ::  | ::||       :   |  |

Db        677 -----EILPIEFLTAYITYAKTKVHPVLTPAAGKALSDAYVNMRKLGDDIRSSDRRITAT 731

Qy        753 VRHIESM 759
                  | :|||
Db        732 TRQLESM 738


RESULT 9
US-10-171-311-129
; Sequence 129, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Monahan, John
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Hoersh, Sebastian
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF CERVICAL CANCER
;  FILE REFERENCE: MRI-035
;  CURRENT APPLICATION NUMBER: US/10/171,311
;  CURRENT FILING DATE:  2002-06-12
;  PRIOR APPLICATION NUMBER: US 60/298,159
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,155
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/335,936
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 129
;   LENGTH: 821
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-129

  Query Match             18.0%;  Score 835;  DB 15;  Length 821;
  Best Local Similarity   27.5%;  Pred. No. 5.4e-64;
  Matches  236;  Conservative  155;  Mismatches  284;  Indels  182;  Gaps   22;

Qy        137 LPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNFLRTHV 196
                | |:|  |||::  |  ::: |  : |                          | :||
Db          3 LAAAAEPGAGSQHLEVRDEVAEKCQKL----------------------FLDFLEEFQ 38

Qy        197 DSHGHNVFKERISDMCKENRESLVVNYEDLAAREHVLAYFLPEA-------PAELLQIFD 249
                | |   : :   :: :  | :|||:: ||      |:  : |           |: |
Db         39 SSDGEIKYLQLAEELIRPERNTLVVSFVDLEQFNQQLSTTIQEEFYRVYPYLCRALKTFV 98

Qy        250 EAALEVVLAMYPKYDRITNHIHVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLP 309
              :   |: ||           :|    ||   ::| |    :  | | || |    | |
Db         99 KDRKEIPLA---------KDFYVAFQDLPTRHKIRELTSSRIGLLTRISGQVVRTHPVHP 149

Qy        310 QLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC--PECQSAGPFEVNMEETIYQNYQRIRI 367
              :|    : |  |  |:    : | :  :|  |  | | :   | ::  :: : ::|::||
Db        150 ELVSGTFLCLDCQTVIRD-VEQQFKYTQPNICRNPVCANRRRFLLDTNKSRFVDFQKVRI 208

Qy        368 QESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIY--------------HNNYDG 413
              ||:  ::    :||| : || |: |:|  |||: : ||                  :
Db        209 QETQAELPRGSIPRSLEVILRAEAVESAQAGDKCDFTGTLIVVPDVSKLSTPGARAETNS 268

Qy        414 SLNTANGFPV---------------FATVILANHVAKKDNKVAVGELTDEDV-------- 450
               ::  :|:                 :  | ||  ||  : :    || ||:
Db        269 RVSGVDGYETEGIRGLRALGVRDLSYRLVFLACCVAPTNPRFGGKELRDEEQTAESIKNQ 328

Qy        451 ------KMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRG 504
                   : :  :|:|: :    :  |: |:|:|::::|||  | |||| ||  |    :||
Db        329 MTVKEWEKVFEMSQDKNLYHNLCTSLFPTIHGNDEVKRGVLLMLFGGVPKTTGEGTSLRG 388

Qy        505 DINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGA 564
```

```
                 |||| : ||| ||||||||::|: | ||::|:|: :||  :|| | |    | |: :||||
Db        389 DINVCIVGDPSTAKSQFLKHVEEFSPRAVYTSGKASSAAGLTAAVVRDEESHEFVIEAGA 448

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGG 624
              |:||| ||| |||||||: :|: :||||||||:|||:|||:  :| || :::|||||| |
Db        449 LMLADNGVCCIDEFDKMDVRDQVAIHEAMEQQTISITKAGVKATLNARTSILAAANPISG 508

Qy        625 RYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEG 684
               || | :  :|::|: ||:||||:  :: |  : | |  :|| :|  |
Db        509 HYDRSKSLKQNINLSAPIMSRFDLFFILVDECNEVTDYAIARRIVDLH------------ 556

Qy        685 LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK--- 741
                  |  | ::   | ::      :::|:::|::   ||:::  :| : : |  ||:
Db        557 ----SRIEESIDRVYSLDD-----IRRYLLFARQ-FKPKISKESEDFIVEQYKHLRQRDG 606

Qy        742 ESMATGSIPITVRHIESMSH--GGGPRAHPSAGLCDRRRRQH------------------ 781
                :   |  |||| :|||        | |     ||  : :|
Db        607 SGVTKSSWRITVRQLESMIRLSEAMARMH----CCDEVQPKHVKEAFRLLNKSIIRVETP 662

Qy        782 ----------------------GH---PRDAGELHRHTEVQRHRSMRKTFAR--YLSFRR 814
                                   ||   |     :: : |     | |   |  :  : |
Db        663 DVNLDQEEEIQMEVDEGAGGINGHADSPAPVNGINGYNEDINQESAPKASLRLGFSEYCR 722

Qy        815 DNNELLLFI-----------LKQLVAEQVTYQRNRFGAQQDTIEVPEKDLVDKARQIN-- 861
               :| ::| :           ||:  :| | :       :: |:    |::|::| | |
Db        723 ISNLIVLHLRKVEEEEDESALKR--SELVNWYLKEIESEIDS----EEELINKKRIIEKV 776

Qy        862 IHNLSAFYDSELFRMNK 878
               || |:  ||  |  : :
Db        777 IHRLT-HYDHVLIELTQ 792


RESULT 10
US-09-828-062-8
; Sequence 8, Application US/09828062
; Publication No. US20030097675A1
; GENERAL INFORMATION:
;  APPLICANT: COSTA E SILVA, OSWALDO DA
;  APPLICANT:  BOHNERT, HANS J.
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  CHEN, ROUYING
;  APPLICANT:  SARRIA-MILLAN, RODRIGO
;  TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
;  TITLE OF INVENTION:  IN PLANTS
;  FILE REFERENCE: 16313-0031
;  CURRENT APPLICATION NUMBER: US/09/828,062
;  CURRENT FILING DATE:  2001-08-20
;  PRIOR APPLICATION NUMBER: 60/196,001
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 901
;   TYPE: PRT
;   ORGANISM: Physcomitrella patens
US-09-828-062-8

  Query Match             18.0%;  Score 835;  DB 11;  Length 901;
  Best Local Similarity   27.5%;  Pred. No. 6.3e-64;
  Matches  247;  Conservative  151;  Mismatches  348;  Indels  152;  Gaps   24;

Qy          2 ASSP-AQRRRGNDPLTSSPGRSSRRTDAL--TSSPGRDL------PPFEDESEGLLGT-- 50
              |:|| | ||      ||:  |  | ||:     |  | |       |::  : |  ||
Db         38 AASPVAGRRAVRQTPTSAVRRRGRETDSARRRRSRSRSLGNSVYSSPYDAGTPGTPGTPV 97

Qy         51 EGPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELTASRREAADGPCGT 110
                |:     |  :   |    :|  |:                    | |
Db         98 ATPVYATPVGTPM---GTPSFHRGTPQY-------------------KQRSE-------- 127

Qy        111 VTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHS 170
                     ||  : |  :   ||::|  |   :    : :   | :  || |  : | |
Db        128 -------LG--SQGKPLHRRRRSQSREPGHRSP---SREPSADGRPSESAEPDDTLGGEY 175

Qy        171 VREWVSMAGPRL---EIHHRFKNFLRTHVDSHGHNV---FKERISDMCKENRESLVVNYE 224
                  |    |  :   ::    : ||  :  |  |::   : : | :  :   ::| ::
Db        176 AYVW----GTNVNIPDVLRAIRRFLHNY-RSSAHDLNSKYIQIIEETVEREEDTLNIDMS 230

Qy        225 DLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNHIHVRISHLPLVEELR 284
              |:   :  |   :  | ::: : |    ||  :: | :::   ||  |  :|     :|
Db        231 DIYDHDPDLYAKIVRYPLDIIPLLDTECQEVATSLLPTFEK---HIEARPFNLKASVHMR 287

Qy        285 SLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPGSC--P 342
               |    :::|:   |:|  |: ::|::      :  | |           :  :  :|  |  |
Db        288 ELNPSDIDKLVSVKGMVIRCSSIIPEIKGAFFKCLVCGHSPPLVTVVKGRVEEPTRCEKP 347

Qy        343 ECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLADLVDSCNAGDEIE 402
              || :     :       : | | :|:||:|  :     | :    |   :||:   || ||
Db        348 ECAARNAMSLIHNRCTFANKQIVRLQETPDAIPEGETPHTVSMCLYNTMVDAVKPGDRIE 407

Qy        403 LTGIYHN---NYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELTDED---------- 449
              :||::         :  |     :: | |   || | |     | |  ||
Db        408 VTGVFKAMAVRVGPNQRTLRA--LYKTYIDCVHVKKSDR----GRLQTEDPMEMDKENDM 461

Qy        450 -----------------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
                              :: :  |||   | :::  |:||||:  ||||:|    ||||:
Db        462 YAGYHESDTSEAANEAKIQKLKELSKLPGIYDRLSRSLAPSIWELEDIKKGLLCQLFGGK 521

Qy        493 PKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRH 552
              |         |||||||| |||||:||| |:|: |:: | |:|:|:|:|||| :|||| :
Db        522 AKKIPSGASFRGDINVLLVGDPGTSKSQLLQYVHKIAPRGIYTSGRGSSAVGLTAYVTKD 581

Qy        553 PVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQAR 612
              | :||  ||:|||||:|||:| |||||||:|  |: :|| ||||::|::| ||: || ||
Db        582 PETRETVLESGALVLSDRGICCIDEFDKMSDNARSMLHEVMEQQTVSVAKGGIIASLNAR 641

Qy        613 CTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSH 672
               :|:| |||  | ||:  |:  :|: |   ::||||:: :: |  |   |  ||| :|  |
Db        642 TSVLACANPSGSRYNARLSVIDNIQLPPTLLSRFDLIYLMLDKPDEQNDRRLARHLVALH 701

Qy        673 VRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKV 732
                ::  :|:                    : |  : |  || ||:: ||| |:    : :
Db        702 YENYEVSKQ-------------------DALDLQTLTAYITYARQHVHPTLSDEAAEDL 741

Qy        733 AKMYSDLRKESMATGS----IPITVRHIESMSHGGGPRAHPSAGLCDRRRRQHGHPRDAG 788
                 | ::|::    ||    |  | | :|||          |  |   |  :     ||
Db        742 INGYVEMRQKGNFPGSSKKVITATPRQLESM-------IRISEALARMRFSEVVEKVDAA 794

Qy        789 ELHRHTEVQRHRSMRKTFARYLS-------------FRRDNNELLLFILKQLVAEQVT 833
              |  |  :|   :|        :                || |   ||  |::|:|::::
Db        795 EAVRLLDVALQQSATDHATGTIDMDLITTGVSASERIRRAN---LLAALRELIADKIS 849


RESULT 11
US-10-128-714-3399
; Sequence 3399, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
```

```
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3399
;   LENGTH: 718
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-3399

  Query Match             17.5%;  Score 810;  DB 15;  Length 718;
  Best Local Similarity   33.4%;  Pred. No. 6.7e-62;
  Matches  228;  Conservative  106;  Mismatches  214;  Indels  134;  Gaps   23;

Qy        131 RRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKN 190
              |||   |   | |  | |:                                  | :|  | :
Db          3 RRTPYTLSVLAPSTDGAEES----------------------------RTQIQARLRE 32

Qy        191 F-LRTHVDSHGHNVFKERISDMCKENRESLVVNY------EDLAAREHVLAYFLPEAPAE 243
              | |   :|    | |  |  |  ::|   ||  |          | :    ||: |   ||:
Db         33 FVLEFQLD----NAFIYR--DQLRQN--VLVKQYFCDIDIAHLISYNEELAHKLTTEPAD 84

Qy        244 LLQIFDEAALE--VVLAMYPKYDRIT--NH---IHVRISHLPLVEELRSLRQLHLNQLIR 296
              :: :| ||||:      :||    |   :|   :|   ||:     :| |   ::: |:|
Db         85 IIPLF-EAALQQCTQRIVYPSQRDIVLPSHQLLLHSSASHI----SIRDLNATNISHLVR 139

Qy        297 TSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPG---------------- 339
              |:|   : :  : ::|   |  |:        |:|  |: |
Db        140 IPGIVIGASTISSKATVVHIRCKGCD-------HSENIRVEGGFSGLTLPRRCGRQKLPG 192

Qy        340 -----SCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLAD--LV 392
                  ||      |: :  |:  : : | :::||:| :|    |||    :: ||  |
Db        193 EEPGEQCP----LDPYVIAHEKCQFVDQQVLKLQEAPDQVPVGELPR--HVLISADRYLA 246

Qy        393 DSCNAGDEIELTGIYH-----NNYDGSLNTANGFPVFATVILA---NHVAKKDNKVAVGE 444
              :    |    : ||:          ||::   |  |    | ::   :| | | | :
Db        247 NRVVPGSRCTVMGIFSIYSKGGKKDGAVAIRN--PYLRAVGISTDLDHTA-KGNAI---- 299

Qy        445 LTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRG 504
              |:|: :    ||: : : | :  ||||||:|: |||:     | ||  |      |:||
Db        300 FTEEEEQEFLELSRREDLYEALARSIAPSIWGNLDIKKAIVCLLMGGSKKILPDGMKLRG 359

Qy        505 DINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGA 564
              |||||| |||||||| ||: ||||  ||:|:|:|:||  :|| ||| | :||: || ||
Db        360 DINVLLLGDPGTAKSQLLKFTEKVSPIAIYTSGKGSSAAGLTASVQRDPQTREFYLEGGA 419

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGG 624
              :|||| ||  ||||||| |:|| :|||||||:|||:|||| | | :| :|:||||| |
Db        420 MVLADGGVVCIDEFDKMRDEDRVAIHEAMEQQTISIAKAGITTILNSRTSVLAAANPIFG 479

Qy        625 RYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEG 684
              |||   |  ||:|    |:||||:: :|||  |  :|| :|| |:| |:|      :: |
Db        480 RYDDLKTPGENIDFQTTILSRFDMIFIVRDDHDRNRDENIARHVMGVHMGGRGIEEQVEA 539

Qy        685 LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK--- 741
                              :| | :|:|| | : |  |:|:    :|::  :  :||
Db        540 -----------------EIPLEKMKRYISYCRSRCAPRLSPEAAEKLSSHFVSIRKQVH 581

Qy        742 ----ESMATGSIPITVRHIESM 759
                ::    ||||||| :|::
Db        582 RAEMDANTRSSIPITVRQLEAI 603


RESULT 12
US-10-128-714-8399
; Sequence 8399, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8399
;   LENGTH: 713
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-8399

  Query Match             17.4%;  Score 804;  DB 15;  Length 713;
  Best Local Similarity   33.3%;  Pred. No. 2.2e-61;
  Matches  227;  Conservative  106;  Mismatches  215;  Indels  134;  Gaps   23;

Qy        131 RRTRSALPASAASGAGTEDGEEDEQMIESIENLEDLKGHSVREWVSMAGPRLEIHHRFKN 190
              |||   |   | |  | |:                                  | :|  | :
Db          3 RRTPYTLSVLAPSTDGAEES----------------------------RTQIQARLRE 32

Qy        191 F-LRTHVDSHGHNVFKERISDMCKENRESLVVNY------EDLAAREHVLAYFLPEAPAE 243
              | |   :|    | |  |  |  ::|   ||  |          | :    ||: |   ||:
Db         33 FVLEFQLD----NAF--IYGDQLRQN--VLVKQYFCDIDIAHLISYNEELAHKLTTEPAD 84

Qy        244 LLQIFDEAALE--VVLAMYPKYDRIT--NH---IHVRISHLPLVEELRSLRQLHLNQLIR 296
              :: :| ||||:      :||    |   :|   :|   ||:     :| |   ::: |:|
Db         85 IIPLF-EAALQQCTQRIVYPSQRDIVLPSHQLLLHSSASHI----SIRDLNATNISHLVR 139

Qy        297 TSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQSQNQEVKPG---------------- 339
              |:|   : :  : ::|   |  |:        |:|  |: |
Db        140 IPGIVIGASTISSKATVVHIRCKGCD-------HSENIRVEGGFSGLTLPRRCGRQKLPG 192

Qy        340 -----SCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLAD--LV 392
                  ||      |: :  |:  : : | :::||:| :|    |||    :: ||  |
Db        193 EEPGEQCP----LDPYVIAHEKCQFVDQQVLKLQEAPDQVPVGELPR--HVLISADRYLA 246

Qy        393 DSCNAGDEIELTGIYH-----NNYDGSLNTANGFPVFATVILA---NHVAKKDNKVAVGE 444
              :    |    : ||:          ||::   |  |    | ::   :| | | | :
Db        247 NRVVPGSRCTVMGIFSIYSKGGKKDGAVAIRN--PYLRAVGISTDLDHTA-KGNAI---- 299

Qy        445 LTDEDVKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGKHKVRG 504
              |:|: :    ||: : : | :  ||||||:|: |||:     | ||  |      |:||
Db        300 FTEEEEQEFLELSRREDLYEALARSIAPSIWGNLDIKKAIVCLLMGGSKKILPDGMKLRG 359

Qy        505 DINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGA 564
              |||||| |||||||| ||: ||||  ||:|:|:|:||  :|| ||| | :||: || ||
Db        360 DINVLLLGDPGTAKSQLLKFTEKVSPIAIYTSGKGSSAAGLTASVQRDPQTREFYLEGGA 419

Qy        565 LVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAAANPIGG 624
              :|||| ||  ||||||| |:|| :|||||||:|||:|||| | | :| :|:||||| |
Db        420 MVLADGGVVCIDEFDKMRDEDRVAIHEAMEQQTISIAKAGITTILNSRTSVLAAANPIFG 479

Qy        625 RYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPSNKEEEG 684
              |||   |  ||:|    |:||||:: :|||  |  :|| :|| |:| |:|      :: |
Db        480 RYDDLKTPGENIDFQTTILSRFDMIFIVRDDHDRNRDENIARHVMGVHMGGRGIEEQVEA 539

Qy        685 LANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRK--- 741
                              :| | :|:|| | : |  |:|:    :|::  :  :||
Db        540 -----------------EIPLEKMKRYISYCRGRCAPRLSPEAAEKLSSHFVSIRKQVH 581

Qy        742 ----ESMATGSIPITVRHIESM 759
```

```
Db        582 RAEMDANTRSSIPITVRQLEAI 603


RESULT 13
US-09-925-300-1053
; Sequence 1053, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1053
;   LENGTH: 724
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (87)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (680)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1053

  Query Match             16.9%;  Score 783;  DB 10;  Length 724;
  Best Local Similarity   31.5%;  Pred. No. 1.6e-59;
  Matches  224;  Conservative  113;  Mismatches  237;  Indels  138;  Gaps   24;

Qy          3 SSPAQRRRGND--------PLTSSPG---RSSRRTDALTSSPGRDLPPFEDESEGLLGTE 51
Db         52 SSPSQRRRGEDSTSTGELQPMPTSPGVDLQSPAAQXVLFSSP------------------ 93

Qy         52 GPLEEEEDGEELIGDGMERDYRAIPELDAYEAEGLALDDEDVEELT----ASRREAA--D 105
Db         94 ------------------------PQM---HSSAIPLDFDVSSPLTYGTPSSRVEGTPRS 126

Qy        106 GPCGTVTGSWPGLGACAVGSCMTAMRRTRSALPASAASGAGTEDGEEDEQMIESIENLED 165
Db        127 GVRGTPVRQRPDLGSAQKGLQVDLQ----------SDGAAAEDIVASEQSL-------- 167

Qy        166 LKGHSVREW---VSMAGPRLEIHHRFKNFLRTHVD---SHGHNV--------FKERISDM 211
Db        168 --GQKLVIWGTDVNVAACK----ENFQRFLQRFIDPLAKEEENVGIDITEPLYMQRLGEI 221

Qy        212 CKENRESLVVNYEDLAAREHVLAYFLPEAPAELLQIFDEAALEVVLAMYPKYDRITNH-I 270
Db        222 NVIGEPFLNVNCEHIKSFDKNLYRQLISYPQEVIPTFDMAVNEIFFDRYP--DSILEHQI 279

Qy        271 HVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 330
Db        280 QVRPFNALKTKNMRNLNPEDIDQLITISGMVIRTSQLIPEMQEAFFQCQVCAHTT-RVEM 338

Qy        331 SQNQEVKPGSCPECQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILLA- 389
Db        339 DRGRIAEPSVCGRCHTTHSMALIHNRSLFSDKQMIKLQESPEDMPAGQTPHT--VILFAH 396

Qy        390 -DLVDSCNAGDEIELTGIYH------NNYDGSLNTANGFPVFATVILANHVAKKDNKVAV 442
Db        397 NDLVDKVQPGDRVNVTGIYRAVPIRVNP-----RVSNVKSVYKTHIDVIHYRKTDAKRLH 451

Qy        443 GELTDED----------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGE 492
Db        452 G--LDEEAEQKLFSEKRVELLKELSRKPDIYERLASALAPSIYEHEDIKKGILLQLFGGT 509

Qy        493 PK--NPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQ 550
Db        510 RKDFSHTGRGKFRAEINILLCGDPGTSKSQLLQYVYNLVPRGQYTSGKGSSAVGLTAYVM 569

Qy        551 RHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQ 610
Db        570 KDPETRQLVLQTGALVLSDNGICCIDEFDKMNESTRSVLHEVMEQQTLSIAKAGIICQLN 629

Qy        611 ARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDE 662
Db        630 ARTSVLAAANPIESQWNPKKTTIENIQLPHTLLSRFDLIFLM---LDP-QDE 677


RESULT 14
US-09-925-301-1382
; Sequence 1382, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1382
;   LENGTH: 766
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (13)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (17)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (123)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1382

  Query Match             16.8%;  Score 778;  DB 9;  Length 766;
  Best Local Similarity   30.9%;  Pred. No. 4.9e-59;
  Matches  197;  Conservative  116;  Mismatches  193;  Indels  132;  Gaps   17;

Qy        187 RFKNFLRTHV--DSHGHNVFK--ERISDMCKENRESLVVNYEDLAAREHVLAYFLPEAPA 242
Db         59 KVKKFLQEFYQDDELGKKQFKYGNQLVRLAHREQVALYVDLDDVAEDDPELVDSICENAR 118

Qy        243 ELLQIFDEAALEVVLAMYPKYD--RITNH--IHVRISHLPLVEE---------------- 282
Db        119 RYAKXFADAVQELL----PQYKEREVVNKDVLDVYIEHRLMMEQRSRDPGMVRSPQNQYP 174

Qy        283 ---------------------LRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKC 321
Db        175 AELMRRFELYFQGPSSNKPRVIREVRADSVGKLVTVRGIVTRVSEVKPKMVVATYTCDQC 234

Qy        322 -----------NFVLGPFCQSQNQEVKPGSCPECQ---SAGPFEVNMEETIYQNYQRIRI 367
Db        235 GAETYQPIQSPTFMPLIMCPSQ----------ECQTNRSGGRLYLQTRGSRFIKFQEMKM 284

Qy        368 QESPGKVAARRLPRSKDAILLADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVF--- 424
Db        285 QEHSDQVPVGNIPRSITVLVEGENTRIAQPGDHVSVTGIF-------------LPILRTG 331

Qy        425 ----------ATVILANHVAK----KDNKVAVGELTDEDVKMITSLSKDQQIGEKIFASI 470
Db        332 FRQVVQGLLSETYLEAHRIVKMNKSEDDESGAGELTREELRQIA----EEDFYEKLAASI 387
```

```
Qy        471 APSIYGHEDIKRGPALALFGGEPKNPGGKHKVRGDINVLLCGDPGTAKSQFLKYIEKVSS 530
Db        388 APEIYGHEDVKKALLLLLVGGVDQSPRGM-KIRGNINICLMGDPGVAKSQLLSYIDRLAP 446

Qy        531 RAIFTTGQGASAVAVTAYVQRHPVSREWTLEAGALVLADRGVCLIDEFDKMNDQDRTSIH 590
Db        447 RSQYTTGRGSSGVGLTAAVLRDSVSGELTLEGGALVLADQGVCCIDEFDKMAEADRTAIH 506

Qy        591 EAMEQQSISISKAGIVTSLQARCTVIAAANPIGGRYDPSLTFSENVDLTEPIISRFDILC 650
Db        507 EVMEQQTISIAKAGILTTLNARCSILAAANPAYGRYNPRRSLEQNIQLPAALLSRFDLLW 566

Qy        651 VVRDTVDPVQDEMLARFV--VGSHVRHHPSNKEEEGLANGSAAEPAMPNTYGVEPLPQEV 708
Db        567 LIQDRPDRDNDLRLAQHITYVHQHSRQPPSQ--------------------FEPLDMKL 605

Qy        709 LKKYIIYAKERVHPKLNQMDQDKVAKMYSDLRKESMAT 746
Db        606 MRRYIAMCREK-QPMVPESLADYITAAYVEMRREAWAS 642
```

RESULT 15
US-10-128-714-3372
; Sequence 3372, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3372
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3372

Query Match 16.8%; Score 776; DB 15; Length 921;
Best Local Similarity 26.3%; Pred. No. 1e-58;
Matches 241; Conservative 152; Mismatches 281; Indels 242; Gaps 29;

```
Qy        138 PASAASGAGTE--DGEEDEQMI----ESIENLEDLKGHSVREWVSMAGPRLEIHHRFKNF 191
Db         36 PPSESNGPLSEIHGGFPDDQVVNPSRSTVSRLRNPYLPGPPPVVDLAGEK--VQQAFEEL 93

Qy        192 LRTHVDSH------------GHNVFKERISDMCKENRESLVVNYEDLAA-REHVLA---- 234
Db         94 LETYVEEPSLSAPPPSSEILSDKYYIAQIHGMKKFELSTLYVDFTHLTSLNNQILADAIA 153

Qy        235 --YFLPEAPAELLQIFDEAALEVVLAMY-PKYDRITNH--------------------I 270
Db        154 NQYF-------RFQPFLTRALHNLIAKYEPEY--FVSHRQASSASSQASSSMMAANASLF 204

Qy        271 HVRISHLPLVEELRSLRQLHLNQLIRTSGVVTSCTGVLPQLSMVKYNCNKCNFVLGPFCQ 330
Db        205 SLAFYNLPLVSRLRQLRTAQIGKLLSISGTVTRTSEIRPELSLGTFVCEQCKSVVSNVEQ 264

Qy        331 SQNQEVKPGSCPE--CQSAGPFEVNMEETIYQNYQRIRIQESPGKVAARRLPRSKDAILL 388
Db        265 T-FRYTEPTQCPNDICMNRSGWRLDIGKSTFVDWQKVKLQESSHEIPTGSMPRTMDVILR 323

Qy        389 ADLVDSCNAGDEIELTGIYHNNYDGSLNTANGFPVFATVILANHVAKKDNKVAVGELT-- 446
Db        324 GEMVDRAKAGERCIFTGTLIVIPDVS---QLGLPGVRPEAVRDNGAFRSGDVGGGGVTGL 380

Qy        447 ----------------------------------------------------DED---- 449
Db        381 KSLGVRDLTYRLAFLACMVTPDTTTPGQQSNQQLNGQSHNILASLNQNNEPEVDEDKAQE 440

Qy        450 ----------VKMITSLSKDQQIGEKIFASIAPSIYGHEDIKRGPALALFGGEPKNPGGK 499
Db        441 ALLQSFTPYEVQDLKNLVHSEYIYSRLVDSIAPMIYGHRQIKKGLLLQLIGGVSKHTEQE 500

Qy        500 H-KVRGDINVLLCGDPGTAKSQFLKYIEKVSSRAIFTTGQGASAVAVTAYVQRHPVSREW 558
Db        501 NMQLRGDINICIVGDPSTSKSQFLKYICSLHPRAVYTSGKASSAAGLTASVVKDAETGEF 560

Qy        559 TLEAGALVLADRGVCLIDEFDKMNDQDRTSIHEAMEQQSISISKAGIVTSLQARCTVIAA 618
Db        561 TIEAGALMLAVRGICAIDEFDKMDISDQVAIHEAMEQQTISIAKAGIHTTLNARASILAA 620

Qy        619 ANPIGGRYDPSLTFSENVDLTEPIISRFDILCVVRDTVDPVQDEMLARFVVGSHVRHHPS 678
Db        621 ANPVGGRYNPKATLRANLNFSAPIMSRFDLFFVIRDEPNETVDRNLADHIVNVHM----- 675

Qy        679 NKEEEGLANGSAAEPAMPNTYGVEPLPQEVLKKYIIYAKERVHPKLNQMDQDKVAKMYSD 738
Db        676 NRDE-------AVQP---------ELSTEQLQRYIRFAR-TFRPVFTEEAKALLVEKYKE 718

Qy        739 LRKESMATG----SIPITVRHIESM---------------------------------- 759
Db        719 LRANDAQGGIGRSSYRITVRQLESLIRLSEAVAKANCVEEIVPSFVREAYDLLRQSIVTV 778

Qy        760 -------SHGGGPRAHPSAG-----LCDRRRRQHGHPRDAGELHRHTEVQRHRSMRKTFA 807
Db        779 EKDDVEVDDDEGTHANAGAGEEDHEMADRDREGDSPMRD------HAEARPRQKTKITYD 832

Qy        808 RYLSF-----RR------------DNNELLLFILKQLVAEQVTYQRNRFGAQQDTIEVPE 850
Db        833 KYMKILNLFVRRVSEDESRSGEGVEQEDLIVWYLEQIESE--------LNSEEDLQR--E 882

Qy        851 KDLVDKA--RQINIHN 864
Db        883 RDLAGKVLKRMVKVRS 898
```

Search completed: September 12, 2003, 13:23:14
Job time : 75 secs